1995

Improving information content in voltammetry experiments

Ming Zhou
San Jose State University

IMPROVING INFORMATION CONTENT

IN VOLTAMMETRY EXPERIMENTS

A Thesis

Presented to

The Faculty of the Department of Chemistry

San Jose State University

In Partial Fulfillment

of the Requirements for the degree

Master of Science

by

Ming Zhou

December 1995

APPROVED FOR THE DEPARTMENT OF CHEMISTRY

Dr. Sam P. Perone (Advisor)

Dr. Joseph J. Pesek

Dr. Roy K. Okuda

APPROVED FOR THE UNIVERSITY

ABSTRACT

IMPROVING INFORMATION CONTENT IN VOLTAMMETRY EXPERIMENTS

by Ming Zhou

This work describes an evaluation of information content of the electrochemical technique known as cyclic voltammetry (CV). The specific informational goal in these studies was qualitative identification of redox couples and electrode processes based on voltammetric shape classification only.

The influence of experimental parameters, such as number of cycles, nCy, and scan rate, V on information content was studied. The voltammetric information content was related to the ability to distinguish among several different electrode reaction mechanisms. The dependence of information content on the number and distribution of patterns was studied. And the influence of selected data processing methods on information content was also considered. The most important experimental parameters were scan rate and number of cycles.

Qualitative classification results with a limited data base were sufficiently accurate to warrant further studies with a more extensive data base.

ACKNOWLEDGMENTS

I am deeply indebted to my major advisor, Dr. Sam P. Perone, for continuously encouraging and advising me to fulfill my entire research studies. I would also like to thank him for many important technical discussions and meticulous review of my thesis.

And, I appreciate Dr. Joseph J. Pesek and Dr. Roy K. Okuda for providing me valuable suggestions and help as my committee members.

Finally, my special appreciations are extended to my family and friends for all their love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Previous Work

Interpretation of analytical data is one of the most challenging problems in chemical analysis. The most efficient and desirable approach to chemical analysis would be to maximize the amount of information obtained relevant to the particular analysis problem. Perone and Ham reviewed the measurement and control of information content in electrochemical experiments (1). The information content of experimental data can be quantitated using Shannon's probabilistic theory (2). Measures of information content have been applied to information theory models of structural analysis, qualitative analysis, quantitative analysis, trace analysis and instrumental analysis. Experimental data are acquired in order to achieve some informational goals. An appropriate figure of merit must be defined with which to measure the extent of informational goal achievement. We can measure success in terms of accuracy, precision, detection limit, resolution, etc. to evaluate the attainment of analytical informational goals.

A few electrochemists have examined the effects of various experimental parameters on the qualitative information content of electrochemical data (3-10).

Several authors have evaluated different methods for the quantitative resolution enhancement of overlapped peaks in programmed potential-step voltammetry (PPSV). Informing power is a crude measure of the amount of

information available in a given analytical procedure and is related to data density and resolution (1). In Perone and Boudreau's (3) study, the informing power magnitude is about 10^3 bits for the smallest useful resolution of 30 mV. The addition of a time resolution element increases the informing power for electrochemical methods to about 10^6 and 10^7 bits.

Thomas and Perone applied pattern recognition techniques to the detection of severely overlapped voltammetric signals on both theoretical (4) and experimental data (5). Their results showed that a k-nearest neighbor (KNN) pattern recognition approach was the most appropriate selection and classification method.

Burgard and Perone (6) reported computerized pattern recognition for classification of 29 organic compounds using Cyclic Staircase voltmmetry (CSCV). The classes examined are almost completely separated on the basis of peak potential. However, their study examined only curve shape information. The results suggested that the information content of the electrochemical data base should be increased for more reliable structural classifications.

Byers and Perone (7) presented a new method of feature weighting for pattern recognition by application to several different voltammetry experimental data sets. The method has been optimized for the K-nearest-neighbor classification rule, and it was observed that K-nearest-neighbor distance weighting is able to increase class separation for oddly shaped distributions.

Schachterle and Perone (8) showed that sufficient information is contained in the shape of a CSCV voltammogram to allow classification of an electrode process according to its mechanism by using pattern recognition.

Similar structures can be expected to react by similar mechanisms, so the approach should also be helpful in qualitative identification by electroanalysis.

Byers, Freiser and Perone (9,10) generated CSCV data for 45 organic compounds to characterize their structure and activity by pattern recognition and electroanalysis.

1.2 Objectives of the Present Studies

The long-range objective of this research is to investigate the information content of programmed potential step voltammetry (PPSV), particularly for Cyclic Staircase Voltammetry (CSCV), because this is the investigation tool of choice for exploratory evaluation. The influence of experimental parameters, such as number of cycles nCy, and scan rate V on information content have been studied.

Because this work is the first step in this project, we have chosen to use the less complicated linear sweep voltammetry (LSV) and Cyclic Voltammetry (CV) to examine a limited set of experimental and process parameters. The information content of CV data was evaluated in relation to the identification of different electrode processes described by the overall mechanism such as preceding, following, or catalytic homogeneous chemical reactions (see Section 3.3.1), as well as heterogeneous kinetic factors.

Chapter 2 Theory

2.1 Programmed Potential Step Voltammetry (PPSV)

The most fundamental stationary electrode voltammetric technique is that where a linearly increasing voltage is applied to a working electrode while measuring the current. This technique is referred to as stationary electrode poloarograph (SEP) or linear sweeping voltammetry (LSV) (11). When a triangular waveform is applied, the method is known as cyclic voltammetry (CV), a great deal of theoretical and experimental development are associated with these techniques (11). However, because the discrete versions of these techniques are more useful experimentally, they will be described here in detail.

Programmed Potential Step Voltammetry (PPSV) is a method of electrochemical analysis in which the potential of an electrochemical cell is changed with variable voltage steps or pulses under program control of a digital computer. With modern digital instrumentation, this technique has been modified for incremental voltage steps, such as Staircase Voltammetry (SCV) and Square Wave Voltammetry (SWV). SCV and SWV are the typical examples of PPSV, which offers the advantage that charging current interference is minimized by measuring cell current at times when the faradaic to charging current ratio is large. Since the double-layer charging current has always been recognized as the major background current in potential-step voltammetric studies, it has been common practice to collect the

data at times longer than four or five cell-time constants, when presumably the double-layer charging current has decayed considerably (12).

Theoretically, the total charging current may be considered as the sum of two current components: the charging current due to the application of a potential step and the charging current induced by the flow of faradaic current. Therefore, in potential-step voltammetry, the total measured current consists of three components: the faradaic current, the step charging current, and the induced charging current. Miaw and Perone (12) also explored the possibility of extracting pure faradaic current from the measured signals.

Fratoni and Perone (13,14) demonstrated the effect of induced charging currents in chronoamperometric studies of flash photolytic processes. In their work, the potentials were located on the diffusion plateau; hence the Cottrell equation was used to describe the faradaic i-t behavior.

$$i(t) = n F A D^{1/2} C_i / (\pi t)^{1/2} \qquad (1)$$

where n = number of electrons

F = Faraday

A = effective electrode area

D = diffusion coefficient

C_i = initial concentration of oxidizable species formed by the flash

t = time

The second advantage of PPSV is the analytical flexibility, because changes in applied voltage wave forms can be implemented by modifying software rather than hardware.

Boudreau and Perone (3) described quantitative resolution of overlapped peaks in PPSV. They showed that voltammograms could be resolved for peak separation greater than 30 mV.

2.1.1 Cyclic Staircase Voltammetry (CSCV)

In 1960, Baker (15) defined Staircase Voltammetry (SCV), where a staircase voltage sweep is imposed on a stationary electrode in a solution, and a series of current pulses are obtained. Figure 1 illustrates the staircase potential function (16).

SCV for reversible electrode reactions was presented first in 1965 by Christie and Lingane (16). It was found that the step height ΔE and the step width τ are the very essential parameters in SCV. The step width is obviously related to the step frequency ($f = 1/\tau$); the sweep rate is equal to $\Delta E/\tau$. It was assumed that the current is measured only at the end of each step, and it was shown that the polarogram obtained is identically equal to that predicted from SEP theory when step height $\Delta E \rightarrow 0$ at constant sweep rate, $\Delta E/\tau$. The peak currents should vary linearly with $(1/\tau)^{1/2}$, and approximately linearly with $\Delta E^{1/2}$.

Zipper and Perone investigated SCV theory by including the effect of varied current sampling time on current-voltage wave forms by a computer-compatible approach to stationary electrode polarography (SEP) in 1973 (17).

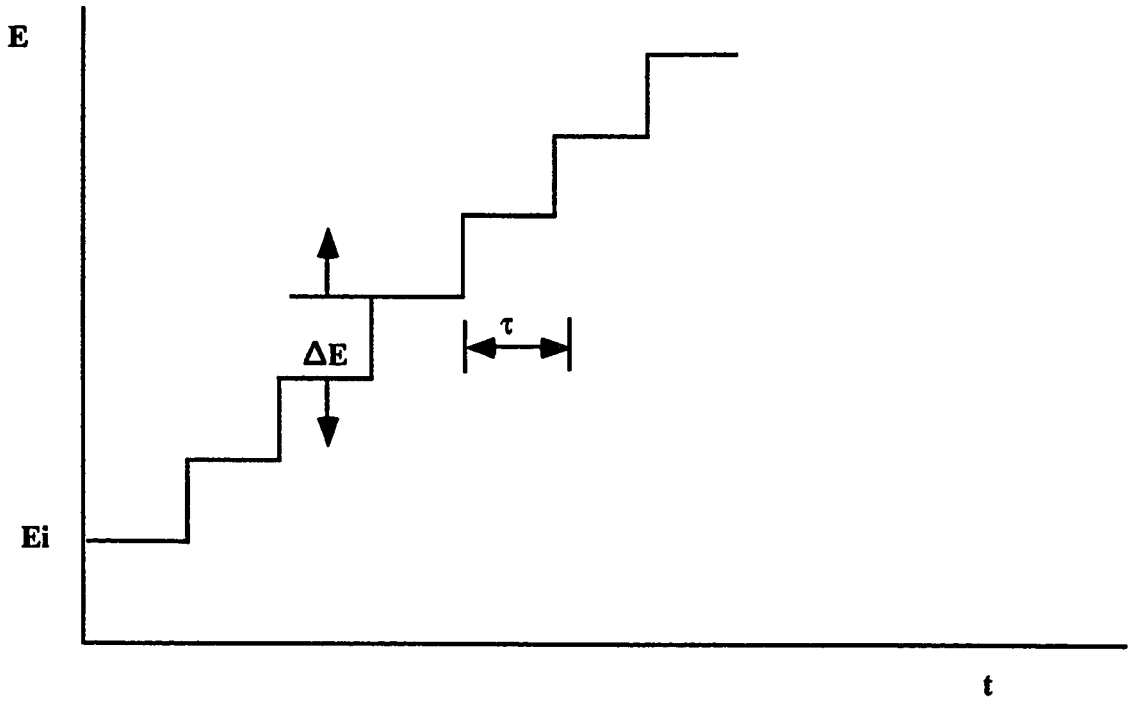

Figure 1. Staircase potential function.

The experimental parameters considered were ΔE, τ, and sampling conditions. Sampling approaches included: varying the sampling time, τ', varying the number of samples, N, averaged on each step, utilizing ensemble averaging with multiple runs for each of the sampling methods mentioned, and least squares digital smoothing.

Cyclic Staircase Voltammetry (CSCV) (18) is applied as shown in figure 2, where the potential is first varied step-wise from the initial potential E_i, in one direction until the switching potential, E_{sw}, is reached, and then is swept back in the same manner. For each step, the step size is designated as ΔE, and the step width, τ. One current datum may be sampled for each step to form a CSCV voltammogram. The time interval between the beginning of the step and the moment the datum is taken is designated as the sampling time, τ'. Theoretically, the shape of a cyclic voltammogram is affected by the ratio of τ'/τ, whereas the absolute magnitude of the voltammogram is directly related to the respective values of τ' and τ. The parameter determined by the ratio of τ' to τ is called the sampling parameter, β, and is defined as $\beta = 1 - (\tau'/\tau)$. β ranges from 0 to 1 depending on the time at which the datum is sampled. If the current value near the end of each step is plotted versus potential, the shape of the polarogram obtained is analogous to Stationary Electrode Pology raphy (SEP). However, like pulse polarography, the measured currents are primarily faradaic and relatively free from charging current. For β close to 0, the data points are taken near the end of the step. For β close to unity, the data points are taken shortly after the beginning of the step. Besides β, the step size ΔE,

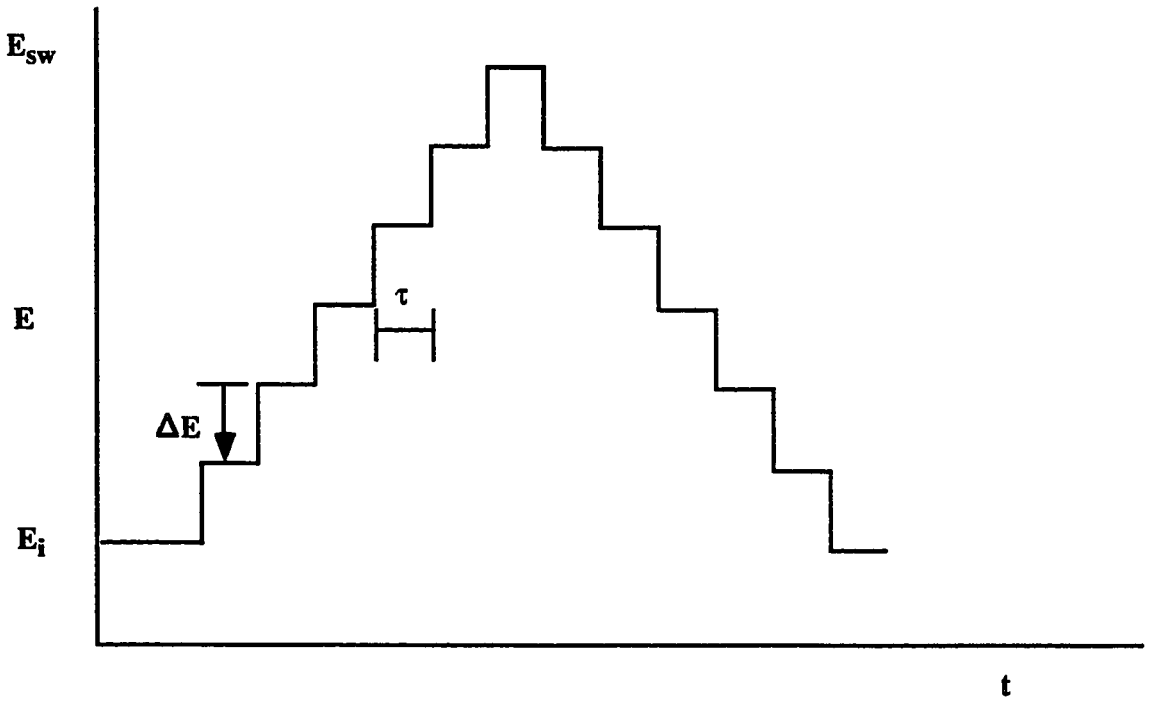

Figure 2. Cyclic staircase potential function.

and the switching potential E_{sw}, are very important in the construction of a cyclic voltammogram.

Bard and Faulkner indicated that Cyclic Staircase Voltammetry (CSCV) is the most effective and versatile electroanalytical technique available for the mechanistic study of redox systems because it enables the electrode potential to be rapidly scanned in search of redox couples (11).

Miaw and Perone (18) conducted CSCV for the reversible electron transfer case experiments to test the effects of potential step size, ΔE, switching potential, E_{sw}, and sampling parameter, β, on peak separation, ΔE_p, $(E_{pa}-E_{pc})$, and peak current ratio, γ, (i_{pa}/i_{pc}).

CSCV had been studied experimentally by Mann (19), who performed several experiments using staircase wave forms, and showed an improved sensitivity over the linear scan technique and largely reduced interference from electrical double layer charging.

The later work by Bilewicz and Osteryoung (20) confirmed the theoretical predictions that for a reversible system Linear Staircase Voltammetry (LSV) and Staircase Voltammetry (SCV) have reasonable agreement, independent of scan rate, provided that current is sampled in SCV at about one-fourth the staircase period ($\beta = 0.75$) and the step height is less than about 8/n mV. For the reversible case, the current in SCV at fixed sampling time is proportional to $(\beta\tau)^{-1/2}$, and the current at $\beta = 0.25$ is the average current over the period τ. It was demonstrated that SCV discriminates against charging current and as a result can be employed over a wider effective time range than LSV.

In early 1960, Nicholson and Shain (21) reported the theory of SEP for both single scan and cyclic triangular wave experiments, which has been extended to electrodes processed involving preceding (CE), following (EC) or catalytic chemical reactions coupled with reversible or irreversible charge transfers. A numerical method was developed for solving the integral equations obtained from the boundary value problems, and extensive data were calculated which permit construction of stationary electrode polarograms from theory. Correlations of kinetic and experimental parameters make possible the development of diagnostic criteria so that unknown systems can be characterized by studying the variation of peak current, half-peak potential, or ratio of anodic to cathodic peak currents as a function of scan rate.

Ferrier and Schroeder studied the quasi-reversible and irreversible electron transfers using SCV (22). Schroeder and Lam studied CE, EC mechanisms using SCV (23). In Ryan's work (24), SCV was used for the study of the catalytic reaction between iron (III)-triethanolamine and hydroxylamine.

2.2 Digital Simulation

The basic concepts of digital simulation of electrochemical phenomena have been discussed in detail by several authors (25-27). Joslin and Pletcher have demonstrated the time-saving virtues of using a non-uniform space grid in digital simulation of electrochemical phenomena (28). They said: "the method has the advantage in simulations of cyclic voltammetry (CV) where

hundreds of iterations are needed in order to present a smooth current-voltage curve or when multiple kinetic processes present a 'stiff' system".

CSCV voltammograms were theoretically generated with digital simulation by Ryan (26), and the use of CSCV in the study of kinetic mechanisms was reported. The validity and usefulness of the technique have been indicated by application to a variety of problems. Perhaps the most difficult aspect of the translation involves calculation of the surface boundary conditions, particularly when there are multiple electron transfers and adsorption phenomena (29, 30).

The outstanding virtue of the digital simulation technique is the ease with which one can translate phenomena into operational mathematical terms (31). The user confronting an unfamiliar CV behavior does not have to carry out or even understand the underlying complex computations and can just consider the mechanism rather than the computation.

The ability to hypothesize reasonable mechanisms can be enhanced by simulating selected systems whose behavior may be characteristic of specific compounds or classes of compounds (11, 28).

Recently, the commercial software DigiSim® 1.0 has become available for computation of theoretical CV responses (32). The DigiSim program is capable of generating voltammograms with a wide range of electrochemical-chemical mechanisms. Because it is the most comprehensive simulation package available, we have chosen to use it for generating theoretical voltammograms. However, the voltammograms correspond to LSV or CV theory, and do not allow study of PPSV parameters like sampling parameter.

Nevertheless, the general effects of scan rate and number of cycles can be studied and results should be comparable to the equivalent PPSV experiments.

Figures 3, 4, 5, 6 show the simulated voltammograms (some of the examples in the study) for different mechanisms and experimental parameters by changing the CV physical/chemical parameters (see section 3.3).

The four simulated voltammograms represent the four different mechanisms: E, EC, CE, ECE (see Section 3.3.1). The basic electrode process, without chemical complication is the E mechanism: $A + ne^- \overset{ks}{\rightleftharpoons} B$, where the heterogeneous rate constant, ks, determines the "reversibility" of the charge transfer step. For low value of ks, the symmetry factor, α, is important in determining the relative potential dependence of the cathodic and anodic steps. α varies between 0 and 1, with 0.5 providing the most symmetric anodic and cathodic step.

The EC mechanism is giving by $A + ne^- \overset{ks}{\rightleftharpoons} B$, $B \underset{kb}{\overset{kf}{\rightleftharpoons}} C$, where the following homogeneous reaction can be characterized by an equilibrium constant K (K = kf/kb), and a rate constant, kf.

The CE mechanism is giving by $A \underset{kb}{\overset{kf}{\rightleftharpoons}} B$, $B + ne^- \overset{ks}{\rightleftharpoons} C$, where the preceding homogeneous reaction can be characterized by an equilibrium constant K (K = kf/kb), and a rate constant, kf.

The ECE mechanism is giving by $A + ne^- \overset{ks}{\rightleftharpoons} B$, $B \underset{kb}{\overset{kf}{\rightleftharpoons}} C$, $C + ne^- \overset{ks}{\rightleftharpoons} D$, where the intermediate homogeneous reaction can be characterized by an equilibrium constant K (K = kf/kb), and a rate constant, kf.

Although many other electrode mechanisms can be defined, only the above four were considered in the study.

For LSV and CV, there is a general relationship between current, experimental parameters, and electrode mechanisms,

$$i = \sum_{M} nFAks\{C_O(0)\exp[(-\alpha nF/RT)(E-E^0)] - C_R(0)\exp[(1-\alpha)(nF/RT)(E-E^0)]\} \quad (2)$$

where M is the number of independent redox couples, A is the electrode area, F is the Faraday, R is the gas constant, T is the temperature, E^0 is the standard potential, and C_O, C_R are concentrations of redox species at the electrode surface which depend on E, T, D, V, K, kf, and mechanism. D = diffusion coefficient of each species, V = scan rate (volts/sec.).

Except for the parameters shown in Table 1 for each voltammogram, the following experimental parameters are the same for all voltammograms in Figures 3-6: T = 298.2K, D = 10^{-5} cm^2/s, electrode area = 1 cm^2 (planar electrode), R_u = 0 (R_u = cell resistance), C (concentration) = 0.001M, E_{sw} = -50 mV, E_i = +1230 mV, E_p = 0.0000mV. All figures except for Figure 4 are 1-cycle voltammograms. For 1-cycle voltammograms, the step size ΔE = 0.0025 V; for 2-cycle voltammograms, the step size ΔE = 0.005 V.

2.3 Information Theory

2.3.1 Shannon's Probabilistic Theory

Information theory studies are based on Shannon's probabilistic model of communication (2,36), which is concerned with the transmission of information through encoding and decoding. The basic principle is that a

Table 1. The experimental parameters for Figures 3,4,5,6.

Parameter / Figure	k_S (cm^2/s)	α	K	k_f (S^{-1})	V (V/s)
Figure 3a	0.005	0.5	\	\	0.05
Figure 3b	0.005	0.5	\	\	1.00
Figure 4	0.05	0.7	50	0.2	0.05
Figure 5	10^6	0.5	0.1	0.2	0.05
Figure 6a	0.0005	0.7	1	0.2	0.05
Figure 6b	0.0005	0.7	1	0.2	1.00

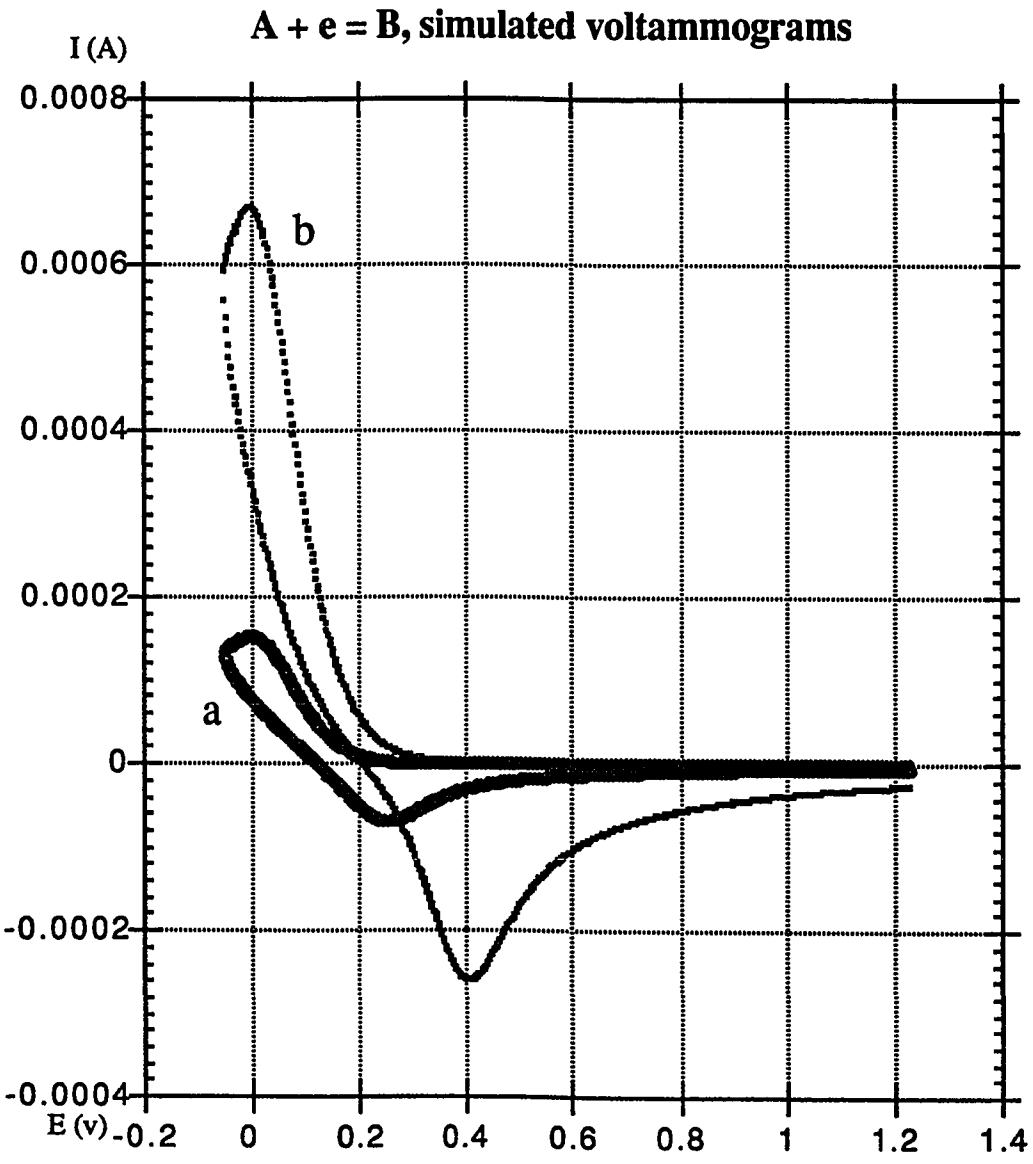

Figure 3. Quasi-reversible electron transfer case, when $k_s = 0.0005, \alpha = 0.5$, 1 cycle, (a): V = 0.05 V/s, (b): V = 1.00 V/s.

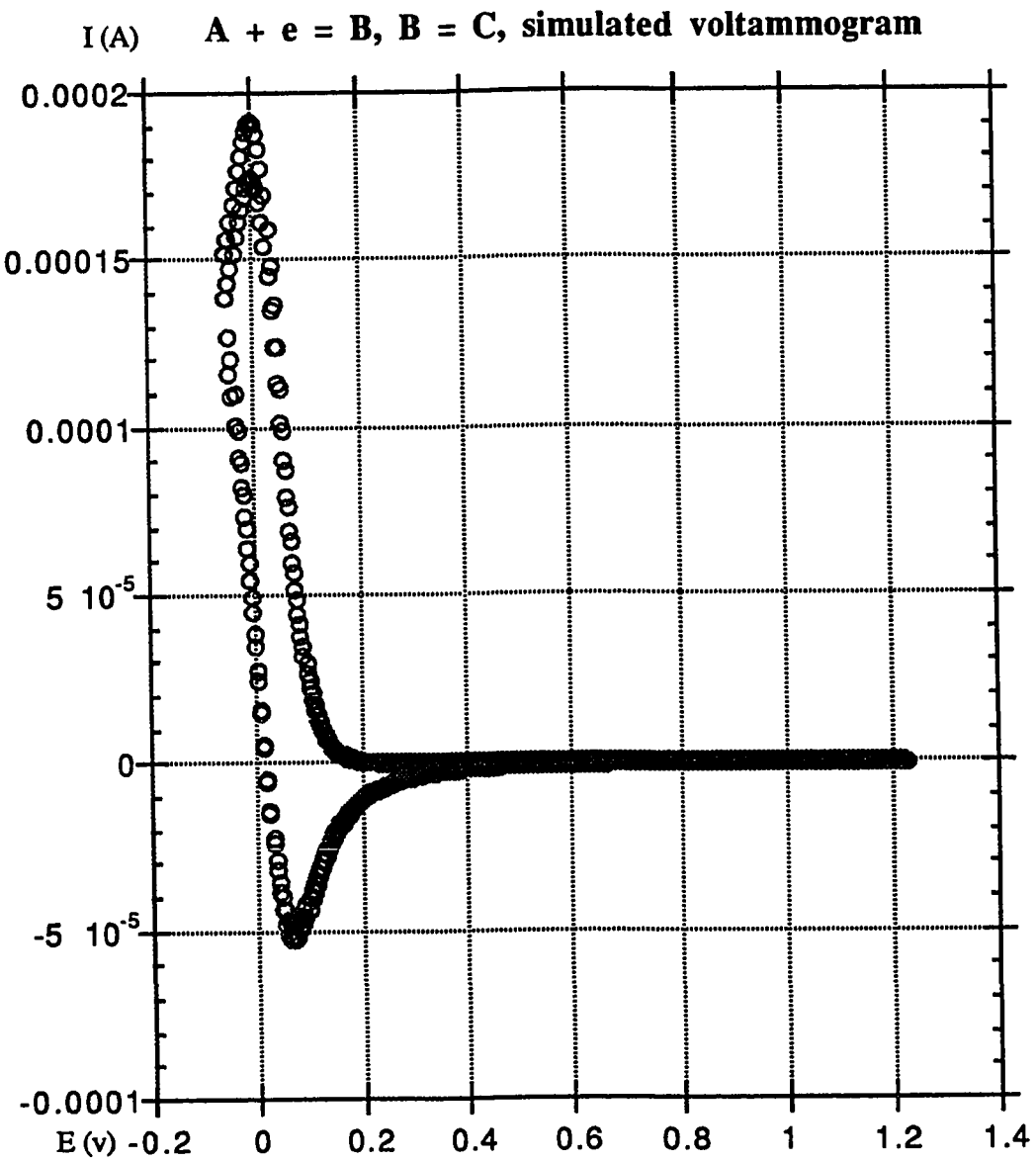

Figure 4. Quasi-reversible electron transfer case, when $k_s = 0.05, \alpha = 0.7$, k = 50, kf = 0.2, 2 cycles, V = 0.05 V/s.

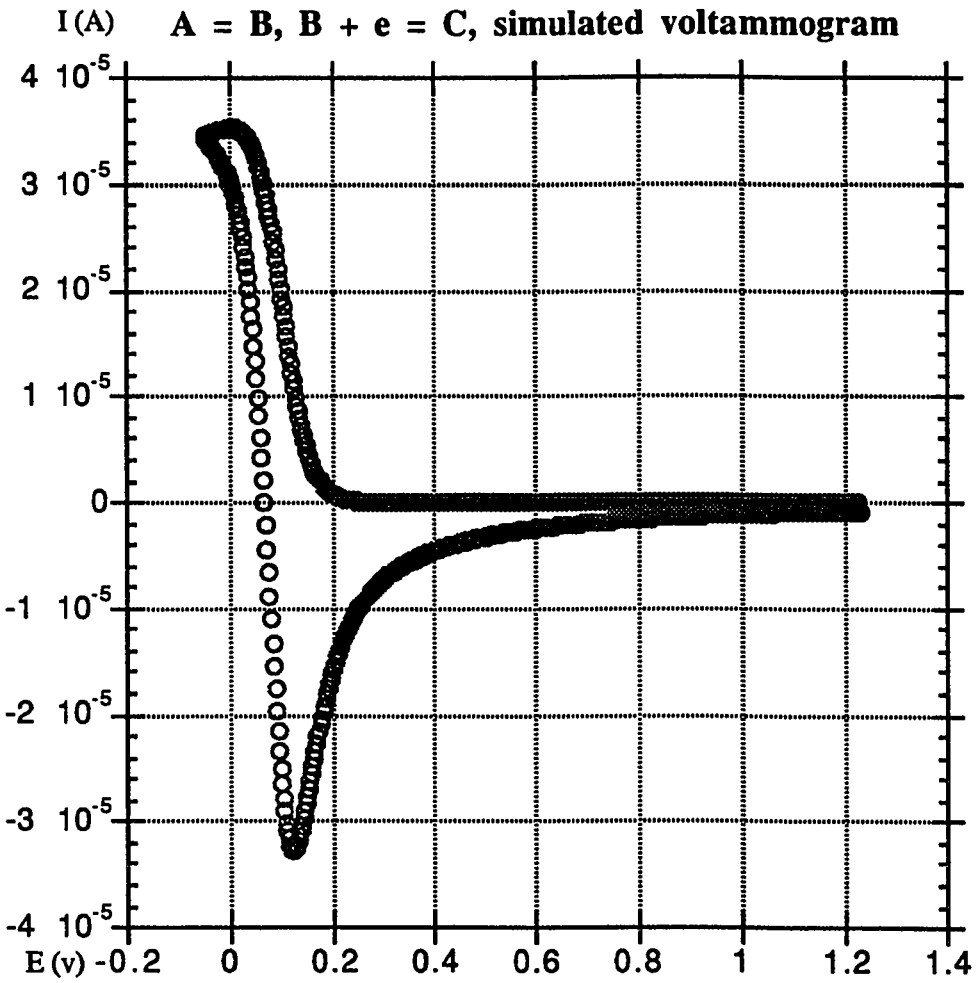

Figure 5. Reversible electron transfer case, when $k_s = 10^6$, $\alpha = 0.5$, k = 0.1, kf = 0.2, V = 0.05 V/s, 1 cycle.

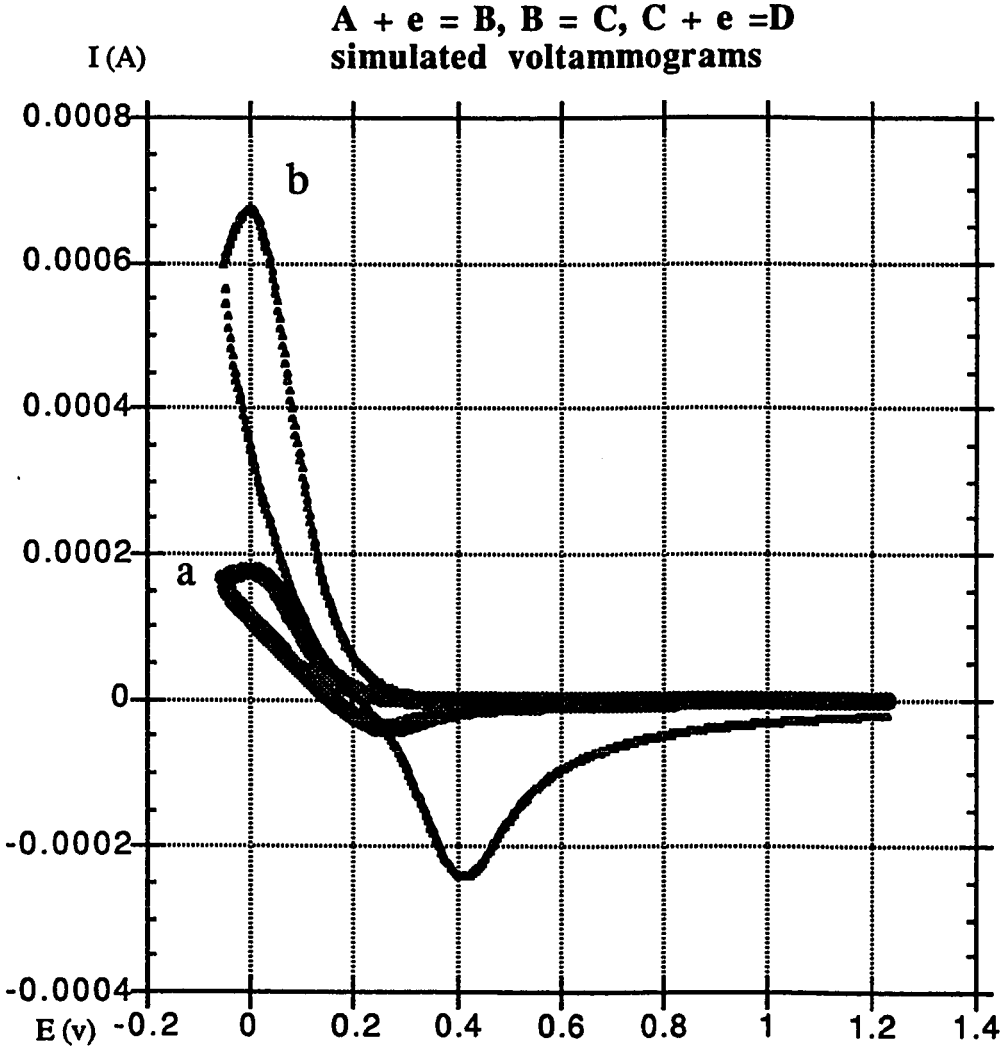

Figure 6. Quasi-reversible electron transfer case, when k_s = 0.0005, α = 0.5, k = 100, kf = 0.2, 1 cycle, (a): V = 0.05 V/s, (b): V = 1.00 V/s.

message with a high probability of random occurrence conveys little information. The most information is conveyed by the message that is least likely to occur spontaneously.

The principle is formalized by the concept of entropy which relates information to uncertainty. Entropy is a quantitative measure of the amount of information supplied by a probabilistic experiment. Shannon's formula, Equation (3) defines the entropy or the average information:

$$H_{av} = I_{av} = -\sum_{i}^{N} p(x_i) \log_2 p(x_i) \qquad (3)$$

Where the symbol $p(x_i)$ is the probability of occurrence of an event, and N is the number of possible events. H is the maximum when events are mutually exclusive and random; then $H_{max} = \log_2 N$.

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly bits, a word suggested by J. W. Tukey (36). A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states is 2^N, and $\log_2 (2^N) = N$. If the base 10 is used as the units , it is called decimal digits. Since

$$\log_2 M = \log_{10} M / \log_{10} 2$$

$$= 3.32 \log_{10} M ,$$

a decimal digit is about 3 1/3 bits.

In analytical work where integration and differentiation are involved the base e is sometimes useful. The resulting units of information will be

called natural units. Changing from the base a to base b merely requires multiplication by $\log_b a$.

2.3.2 Information Content of Chemical Data

In 1985, Perone and Ham (1) reviewed some of the basic concepts in information theory, including history, terminology, entropy, and other information content measures. Chemical applications of information theory which have been reported in the literature including applications to qualitative analysis, quantitative analysis, and structural analysis were also reviewed. Measures of information content and figures of merit for performance evaluations were discussed.

There are four kinds of chemical measurements relating to informational composition (35):

Essential data: a minimum sub-set of the most relevant data.

Redundant data: relevant, but non-essential, confirmatory data.

Informational noise: irrelevant, but still valid data.

Measurement noise: random measurement error.

The analyst is usually only interested in a subset of the data which have been collected. The relevant chemical information is dependent not only upon the informational goals of the problem, but the completely specified procedure as well. The use of information theory in analytical chemistry originates from

the idea that each analysis is a process of obtaining information about the qualitative or quantitative composition of a sample, or information about the structure of the analyzed substance (36). The information content provided by particular analytical procedures or devices is important because it enables us to compare their relative suitability for different applications. Optimization can be performed by employing an information quantity as a response function.

Perone and Ham (1) provided a detailed discussion of the application of information theory to electrochemical experiments and the empirical determination of the information content of electroanalytical data. In Byers, Freiser, and Perone's studies (9, 10), the data set of 45 organic compounds analyzed by using CSCV consisted of 19 nitrobenzenes, 9 nitrodiphenyl ethers, and 17 ortho-hydroxy azo compounds. Of the nitrodiphenyl ethers, 4 were strong herbicides and 5 were either weak or nonherbicides. The informational goals for the problem were the classification of the 45 compounds by their structural type and the classification of the 9 nitrodiphenyl ethers according to herbicidal activity. The figure of merit for the informational goals was the classification accuracy as achieved by k-nearest neighbor analysis. Thus, related information content was determined for various data sets based on their effectiveness for compound classification by pattern recognition.

In this study, qualitative identification of redox species is the information goal for cyclic staircase voltmmetry (CSCV) experiments. And, we have attempted to describe information content quantitatively using Shannon's theory (see Chapter 4).

2.4 Pattern Recognition

The literature on application of pattern recognition approach to chemical data interpretation is already voluminous (35, 37-42). Among these are several which deal with problems that previously appeared to have only extremely difficult, complex, or even unrealizable solutions. Numerous such problems have proved amenable to attack by pattern recognition techniques.

Pattern recognition involves the detection and recognition of regularities among sets of measurements describing objects or events. It is concerned with the investigation of analytical techniques for processing large amounts of data, and extracting useful information to reduce the data.

Pattern recognition is concerned primarily with the description and analysis of measurements taken from a physical system. Data observed (d-dimensional pattern space) are the experimentally determined values used to characterize the objects. A pattern is defined as a d-dimensional vector composed of d calculated independent data items, and can be represented by:

$$P = W_1 X_1 + W_2 X_2 + W_3 X_3 + \text{------} + W_d X_d \qquad (4)$$

X_1, X_2, X_3 ------, X_d = components (measurements) of the pattern vector;

W_1, W_2, W_3, ------, W_d = components of the weight vector;

d = number of dimensions

2.4.1 Classification Method: KNN

Pattern recognition assumes that the characteristic properties or individual identities of objects which distinguish one from another can be determined by examination of pattern vectors associated with each object. Classification involves making a proper decision to assign each pattern to one of the possible pattern classes. The decision is usually based on comparing pattern vectors for unknown objects with those of known objects. It is assumed that pattern vectors for items of the same class occupy the same region of space.

The K-nearest neighbor (KNN) classification is a standard method in pattern recognition and is a very powerful tool for its elegant simplicity. Figure 7 (37) illustrates an example of 2-dimensional KNN. In this method, each pattern can be considered as a point in the N-dimensional feature space.

The distance between any two pattern vectors, i and j, in the N-space is defined by Euclidean geometry, and Euclidean distance calculation for pattern vectors in N-space is shown in Figure 8 (37). Euclidean distance calculation is represented in Equation 5:

$$D_{ij} = \left[\sum_{n=1}^{N} (X_{in} - X_{jn})^2 \right]^{1/2} \qquad (5)$$

D_{ij} = distance between two pattern vectors in N space

i, j = refer to specific pattern vectors i and j

n = index for features

N = total number of features

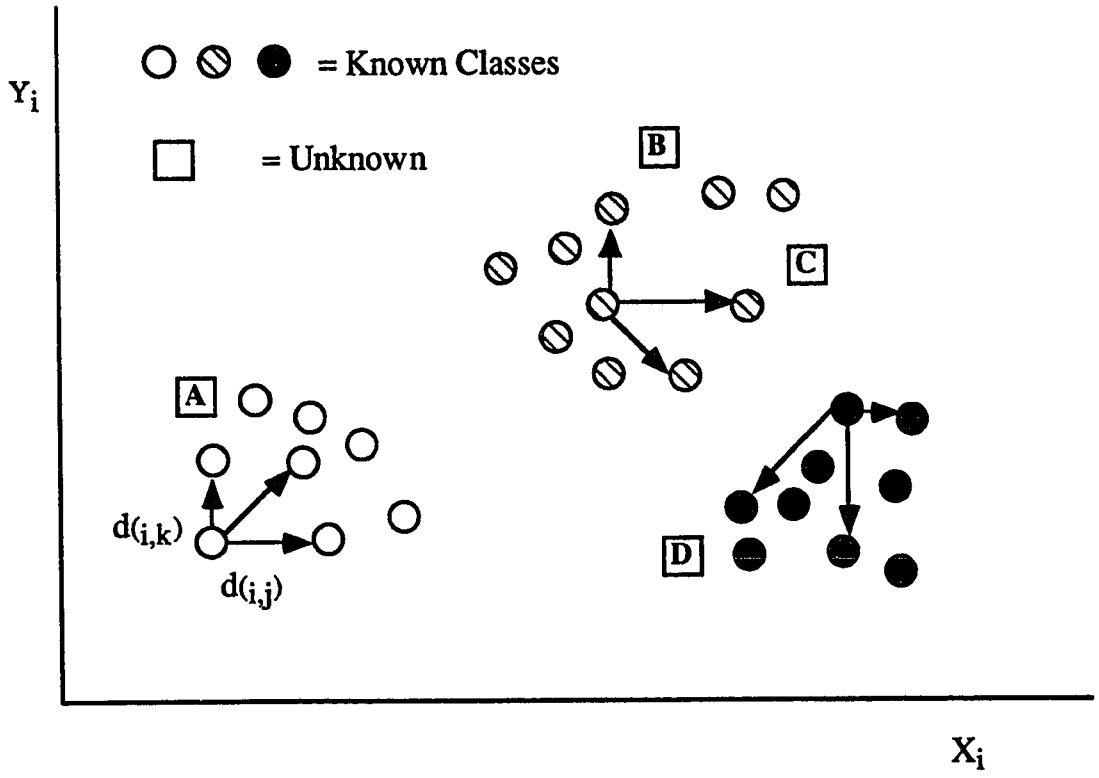

Figure 7. 2-dimensional KNN example.

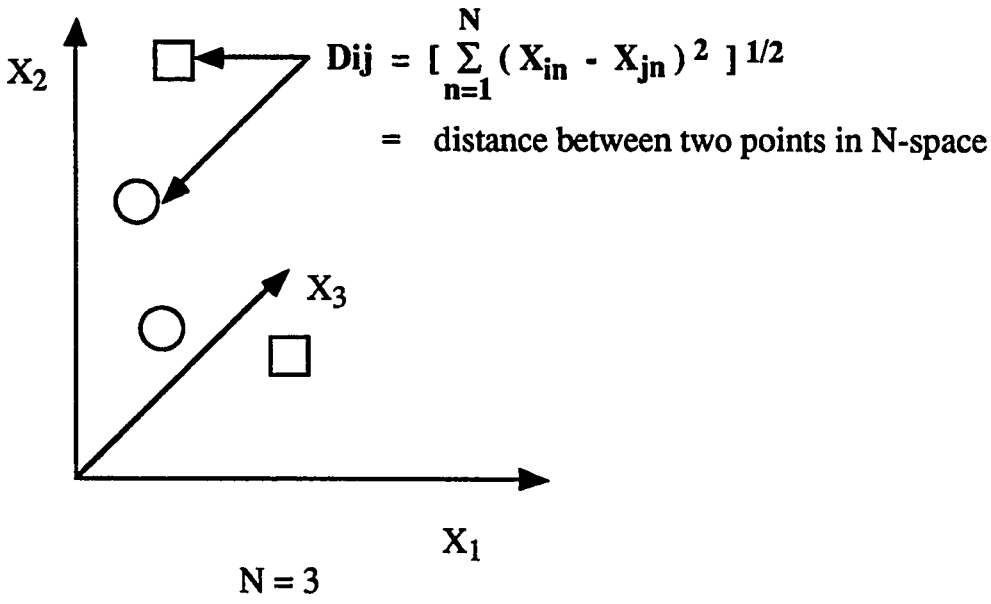

Figure 8. Euclidian distance calculation for pattern vectors in N-space.

X_{in} = value of feature n for pattern i

X_{jn} = value of feature n for pattern j

In order to find out the nearest neighbors for each pattern, it is necessary to compute the distance between each specific pattern and all other pattern vectors in the feature space. In the classification procedure, an unknown pattern is classified according to the majority vote of its K-nearest neighbors. For two class systems, K is an odd number of neighbors which are considered for classification. Usually a small value of K is preferred, and K = 1 is the most common choice, particularly for multi-category classification (greater than 2 classes).

Supervised pattern recognition technique involves identifying an unknown item by including its pattern vectors in a set of pattern vectors for items of known class. Identification is based on the majority class of nearest neighbors by computerized calculation of inter-item distances in feature space. Figure 9 (35) gives an example of a supervised pattern recognition procedure.

2.4.2 Feature Weighting

There are two purposes of feature weighting: it can provide a measure of the discriminating ability of a variable in terms of category separation; it

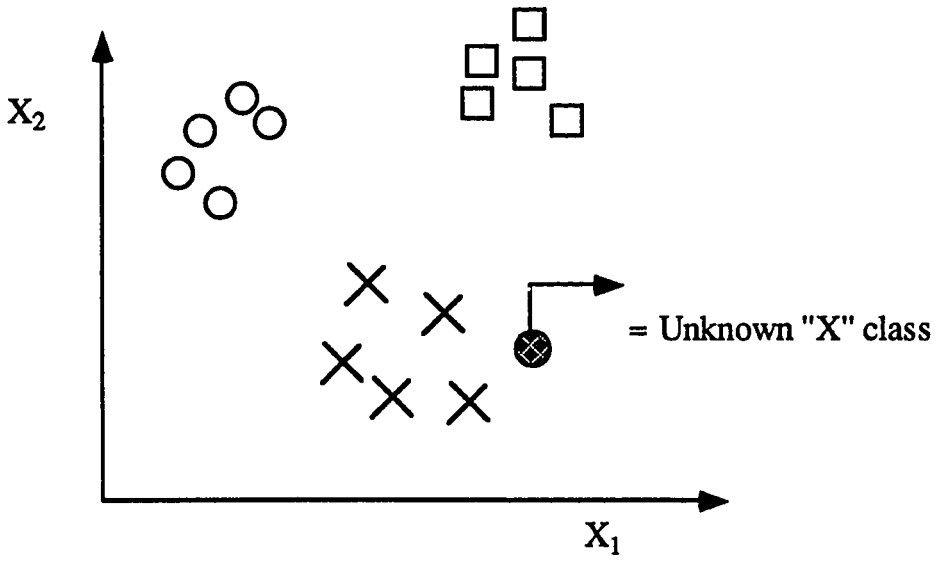

Figure 9. Supervised pattern recognition technique example.

can improve classification results by "stretching", that is, scaling each axis according to its overall feature weight.

The supervised pattern recognition approach to weighting has two main advantages over other methods. First, the interaction between different features can be taken into account when choosing weights with supervised learning. Second, the weights chosen will be optimized for the particular classifier being used.

The supervised pattern recognition approach to weighting chemical features has been shown to be superior to Fisher (39) or Variance weighting (39) when there is partial class overlap in the data being analyzed (7). The superior performance of the supervised pattern recognition approach is largely due to the development of a criterion for stretching or shrinking the axes of a multidimensional feature space which is closely related to the K-nearest neighbor classification rule. Since the K-nearest neighbor rule can produce good classification for oddly shaped class distributions, K-nearest neighbor distance weighting is able to increase class separation for oddly shaped distributions.

Also, it makes classification more insensitive to accidental inclusion of poor features. If a poor feature is entered into the model, its weight will later be adjusted to a value near zero, removing its influence in the classification step.

The procedure of optimum feature weighting for the K-nearest neighbor classifier is that each feature is first auto-scaled over all patterns so that it has a mean of 0 and a standard deviation of 1. This gives all features equal weight to begin with. All features are then multiplied by an arbitrary

scalar, and class predictions are made. An error term is accumulated which is very closely related to the KNN classification accuracy. Feature weights are adjusted in an iterative trial-and-error procedure by increasing the weight for one feature at a time and the training; this procedure continues as long as the error term decays, and stops when the error term increases.

The ultimate goal is to reduce classification error, so the error function used in the process is very important. The error terms used in this work are discussed below (see section 3.5).

2.4.3 Procedures

2.4.3.1 Training, Prediction, Unknown Set

Various Pattern sets are defined as follows:

Training (prototype) set: set of patterns for items of known classes (35).

Prediction set: set of patterns of same classes as training set, known, but not contained in training set, used to test validity of classification principles based on training set (35).

Unknown Set: set of patterns containing same measured parameters as training set, but where actual class information is not available (35).

Pattern recognition technique is usually first carried out on a training set to develop a classifier that recognizes the class membership of these patterns as well as possible. The procedure is called training. The true identity of each pattern is compared to the identity assigned to it in the

classification step; the percentage of correctly classified training set patterns is called recognition rate, also it is the classification accuracy of the classifier.

If the recognition rate is poor, then other measurements or alternative data transformations should be tried until an acceptably high accuracy is achieved, if possible.

Once a sufficiently accurate decision rule has been obtained by the training procedure, we still need to demonstrate that it is reliable, and not fortuitous. This can be done by observing how successful it is at classifying a different set of known patterns, the prediction set. These patterns have essentially the same origins as the training set and represent the same classes included in the training set. This process of applying the trained classification rule to the prediction set is called prediction. The percentage of correctly classified prediction set patterns (which have not been used in training) is called prediction ability. If patterns of the prediction set are also correctly classified with high accuracy, the classification rule can be considered valid and reliable. If the decision rule (classifier) is successful for the prediction set, it can be concluded that the sought for information is indeed contained in the data matrix. Truly unknown patterns may then be analyzed as long as each is collected in the same manner as the training/prediction sets and belongs to one of the classes represented in the training/prediction sets.

Figure 10 is the diagram of pattern classification (35).

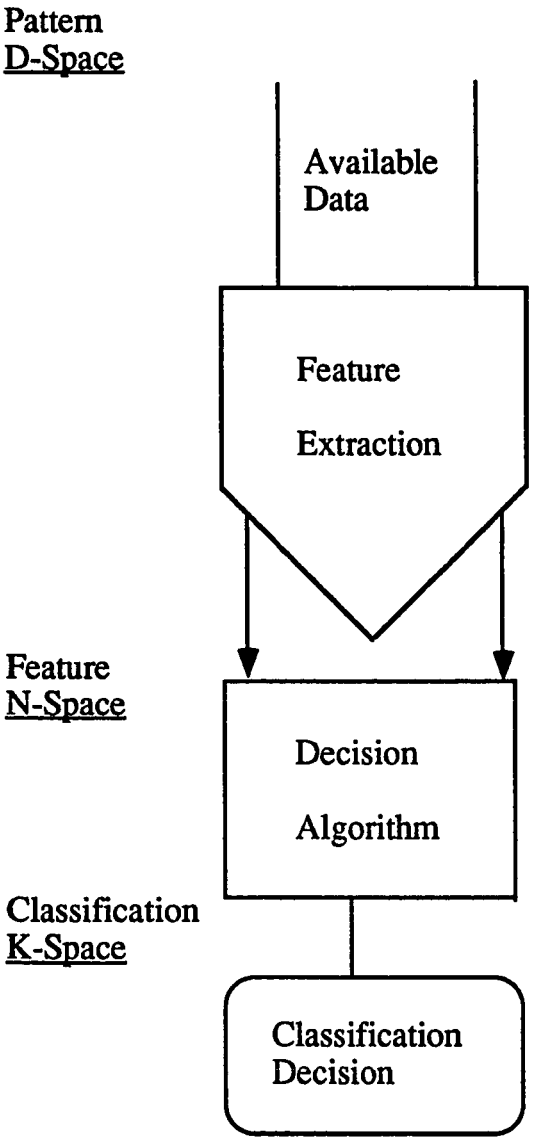

Figure 10. Pattern Classification.

2.4.3.2 Evaluation of Pattern Recognition results

2..4.3.2.1 Classification Accuracy

The following terms describe the classification accuracy for evaluating pattern recognition results.

A = overall accuracy

= (number of correct * 100%)/total patterns in set

= correct *100% / J (6)

A_R = class specific accuracy = C_R * 100% / J_R (7)

$\overline{A}_R$ = average accuracy = $\Sigma A_Z / Z$ (8)

where J = total number of patterns

C_R = the number of correct for class R

J_R = number of patterns of class R

Z = total number of classes

2.4.3.2.2 Nonlinear Mapping Display

The nonlinear mapping technique (44,45) plots n-dimensional data in two dimensions, while attempting to minimize the difference between N-space and corresponding 2-space interpoint distances by iteratively minimizing the mapping error. The mapping error is defined:

$$E = [\sum_{i<j}^{p} d_{ij}]^{-1} \Sigma (d_{ij} - d_{ij}^{*})^2 / d_{ij} \quad (9)$$

where d_{ij} = distance in N-space

d_{ij}^{*} = distance in 2-space

P = total number of patterns

Since the mapping is a 2-dimensional display, it allows visual verification of pattern vector separation by class in N-space. It also allows identification of the nature of classification errors. Nonlinear mapping is extremely useful after initial selection of a small set of features, as it guides the selection of optimum feature subsets.

2.4.3.2.3 Information Content

Perhaps the most meaningful criteria for evaluation of pattern recognition results are measures of information content. In this work, we explore for the first time the use of information theory principle (section 2.3) to construct figures of merit for pattern recognition results. The procedures adopted for this are described later (Section 4.4). Most importantly, this approach can establish quantitative criteria for selecting optimum experimental conditions for achieving specific informational goals.

2.4.4 Feature Selection Techniques

Because the raw data vector may be of large dimension, some reduction of dimensionality is desired to obtain subsequent reliable classification (43).

The function of feature selection is to extract or to measure the most informative characteristics from the input patterns. This step involves the definition of N-dimensional feature space where N < d (N, d definition: see Section 2.4). Features defined from the raw data may include the individual data items as well as transformations and combinations of these data. The extracted features (pattern descriptors comprising the lower dimensional presentation) are supposed to best characterize the distinguishing properties of pattern classes.

Four kinds of feature selection techniques are commonly used for pattern recognition: intuitive, numerical transforms, statistical, and systematic trial-and error. These approaches are discussed below.

2.4.4.1 Intuitive

This method is used directly by imposing the investigator's personal knowledge of existing correlations between system properties or some prior expectation of what data features are most likely to contain information for specific classification problems. It requires some awareness of physical/chemical significance of data and extracted features. The disadvantage of this method is that information content is not obvious for multivariate feature interactions or for previously unexplored classification issues. However, intuition often provides a good starting point for feature selection, followed by multivariate empirical classification studies.

2.4.4.2 Numerical Transforms

Many possible numerical transform techniques might be used to reduce data dimentionality while retaining desired information content (46). Of these, the Fourier Transform has been the most useful for voltammetry data (46), and will be discussed here.

Fourier Transform is a technique for solving problems in linear systems and analyzing periodic data. It is expressed in the following equations:

$$F(\nu) = \int_{-\infty}^{+\infty} f(t)\, e^{-2\pi i t \nu} dt \qquad (10)$$

$F(\nu)$ = frequency domain signal = function of frequency, ν,

$$F(t) = \int_{-\infty}^{+\infty} f(\nu)\, e^{2\pi i t \nu} d\nu \qquad (11)$$

$F(t)$ = time domain signal = function of time, t,

The algorithm which allows the efficient computation of a transform for discrete frequencies (or time intervals) is called the Fast Fourier Transform (FFT) (46). A discrete transform is represented schematically by Equation 12.

$$F(\nu) = a_0 f(\nu_0) + a_1 f(\nu_1) + \ldots\ldots\ldots\ldots\ldots a_n f(\nu_n) \qquad (12)$$

FFT achieves dimensionality reduction and noise reduction when only lower frequency components are selected (35). It also emphasizes frequency content of data, which can be very useful if relevant to informational goals.

Thomas and Perone (4,5) have shown that the coefficients obtained from the discrete fourier transform of voltammetric curves are useful for curve shape analysis. FFT features seem to allow the most reliable and consistent classification of voltammetric doublets and singlets. Features derived from the discrete FFT have been shown to be useful for the enhancement of intrinsic class differences in real voltammetric data. Thus, these features from FFT are apparently good enough for the representation of analytical wave forms for pattern recognition, when intrinsic wave shape information is sought.

Burgard and Perone (6) studied computerized pattern recognition for classification of organic compounds from voltammetric data, and found that curve shape features obtained by FFT are very useful for accurate classification of the voltammograms.

2.4.4.3 Statistical

There are two commonly used statistical methods: correlation analysis and Fisher ratio (35).

Correlation Analysis: correlation analysis (47) involves calculation of the linear correlation coefficients for each possible pair of features in the pattern-feature matrix. The value of the correlation coefficient is within +1.0 to -1.0. The value +1.0 represents highest positive correlation, e.g., identical

numerical data for two sets of features; the value -1.0 stands for highest negative correlation, e.g., equal but opposite sequence of numerical data for two sets of features; the value 0 represents no correlation between two features. When two features are highly correlated, one can be eliminated without loss of statistical information. Thus, feature elimination by correlation analysis can condense all features with mutually large correlation coefficients into a smaller set containing the same statistical information.

However, it is not always inappropriate to have highly correlated features in the pattern recognition data base, especially when the distribution curve is not normal Gaussian. Thus, correlation analysis must be used with caution as a feature elimination tool.

Fisher Ratio: it measures class separability and distribution (39), and it is expressed in Equation 13.

$$(F.R.)_n = (\bar{X}_{Z_1} - \bar{X}_{Z_2}) / (S^2_{z1} + S^2_{z2}) \qquad (13)$$

where $(F.R.)_n$ = Fisher Ratio for feature n

$\bar{X}_{Z_1}$ = an average value of feature n for class Z_1

$\bar{X}_{Z_2}$ = an average value of feature n for class Z_2

S^2_{z1} = a standard deviation value for feature n, Class Z_1

S^2_{z2} = a standard deviation value for feature n, Class Z_2

2.4.4.4 Systematic

There are two commonly used methods for systematic feature elimination: one-dimensional feature performance and systematic trial-and-error (35).

One-dimensional Feature Performance: this procedure involves examining one feature at a time to determine its discriminating ability in one dimension. All features are then ranked in decreasing order of univariate classification accuracy. Only the higher-ranked features are selected for subsequent multivariate pattern classification studies. This procedure achieves a first step towards condensation of features into the smallest useful set. The basic assumption is that features with higher univariate classification accuracy should be the most useful for multivariate classification. This is a reasonable, but not necessarily valid assumption. Nevertheless, the one-dimensional feature selection procedure is useful.

Systematic Trial-and-Error: this feature elimination procedure depends on multivariate classification performance. First, a fixed number of features are applied to a multivariate classification problem, and the classification accuracy is obtained for the complete set. Then, one feature at a time is eliminated to see whether the classification result changes or not. If the result is improved or the same, it means the eliminated feature is not necessary, and it can be rejected; if the result is degraded, this feature should be retained for classification. After the elimination/retention decision is

made, the testing and elimination procedure can be continued until each feature had been evaluated. Usually, only a fraction of the features is retained, and the classification accuracy is improved over the original set. Other features may then be considered, and the procedure can be repeated. Selected features from separate elimination studies can be combined and the elimination procedure applied again, until achieving the best classification accuracy with a minimal set of features.

In the feature elimination process, it is also possible to vary and optimize weights of useful features. Finally, the most useful features together with their optimum weights will give the best classification ability, with a minimum set of features. However, the absolute best set of features may not be found with this procedure unless all possible combinations are considered. With large feature sets, this is not practical, but the sequential feature elimination procedure described here works quite well. Practically, if the sequence in which features are eliminated is varied over several different feature elimination cycles, the consistently useful features are clearly recognized.

Chapter 3 Experimental Approach

3.1 Instrumentation

An IBM clone personal computer (PC) is used for database management and computerized digital simulation (DigiSim® software), fast fourier transform (Microsoft® EXCEL software package), K-nearest neighbor pattern recognition (in-house KNNPRA software package) (7).

DigiSim® 1.0 software runs in the DOS environment, and source code (not provided) is programmed in C++. The C++ compiler has effected a DOS extender which allows DigiSim to utilize as much of the RAM as needed. Although many applications of DigiSim will run within the traditional 640K-byte DOS limit, at least 2 M-bytes of extended RAM is recommended. DigiSim 1.0 will run only with a DOS based configuration. The acceptable configuration requires 2 Mbyte RAM and 20 Mbyte hard disk, on a 386 DX/33 MHz PC, while a 486 DX $\geq$ 25 MHz with 2-M-bytes extended RAM is highly recommended. A typical digital simulation for cyclic staircase voltammetry requires about 12 s on a 486 DX / 66 MHz PC.

In this study, the digital simulation run for one cycle CV, step size ΔE = 0.0025 V, E mechanism is about 11.15 s; EC mechanism is about 22.08 s; CE mechanism is about 22.08 s; ECE mechanism is about 69.15 s. There is no big difference in run time when α, k_s, K, k_f are changed. For two-cycle CV, because the step size ΔE is reduced to half of the single-cycle CV, step size ΔE = 0.005 v, the run time is about the same as the one-cycle.

Fast fourier analysis can be performed under Microsoft® EXCEL 4.0 or 5.0 with data analysis tool package. It takes about 2 minutes for 1024 data points on a 486 DX /25 MHz PC, but 10 s on a 486 DX-2 / 66 MHz PC.

For pattern recognition, the minimum configuration requires 1 M byte RAM and 20 M byte hard disk, with a 4 MHz 286 microprocessor. For some more complicated applications, a similar system with a 12 MHz 386 microprocessor and 8087 math co-processor was preferred. In this study, a typical pattern recognition run requires about one minute on a 486 DX-2 / 66 MHz PC, but it takes more than fifteen to thirty minutes to process class-weighted KNN training with feature optimization; and the run time is substantially longer when number of patterns or features are increased. The software KNNPRA was written in Microsoft compiled BASIC, version 1.1.

3.2 Data Base Management and Data Analysis

After all raw data are generated from DigiSim®, they are contained in a Microsoft® EXCEL spreadsheet data base management system. Rotated cyclic voltammograms (CV), basic statistical computations (averages, variances, distributions, maxima, minima, etc.), fourier feature transform calculation, normalized fourier feature transform data, and associated graphical procedures are conducted using packages contained within Microsoft® EXCEL. Multivariate analysis procedures of K-nearest neighbor pattern recognition (such as KNN training with feature elimination, and KNN training with class weighted training) are conducted with the KNNPRA software .

3.3 Digital Simulations

3.3.1 Experimental Mechanisms and Compound Classes for CV

The purpose of this study was to determine voltammetric information content related to identification of different classes of compounds. To represent voltammetric data for various compounds, voltammograms were simulated for various combination of physical/chemical properties. Each unique set represents a unique compound. Classes of compounds were represented by simulations grouped according to basis simulations of mechanism E, EC, CE, and ECE.

There are four basic mechanisms considered here:

E mechanism: uncomplicated electron transfer,

$$A + e = B$$

EC mechanism: following chemical reaction,

$$A + e = B$$

$$B = C$$

CE mechanism: preceding chemical reaction,

$$A = B$$

$$B + e = C$$

ECE mechanism: combination of following and preceding chemical reaction,

$$A + e = B$$

$$B = C$$

$$C + e = D$$

For pattern recognition studies, each simulated voltammogram obtained with a different combination of physical/chemical properties represents a different "compound" Thus, five classes of compounds have been defined representing different basic electrochemical processes:

Class 1: E mechanism with reversible electron transfer ($k_s >> 10^3$).

Class 2: E mechanism with quasi-reversible electron transfer ($0<k_s<10^3$).

Class 3: EC mechanism with both reversible and quasi-reversible electron transfer.

Class 4: CE mechanism with both reversible and quasi-reversible electron transfer.

Class 5: ECE mechanism with both reversible and quasi-reversible electron transfer.

3.3.2 Experimental Parameters for CSCV

Because this study considered only voltammetric parameters, and the one objective of this study is to determine correlations between information content and experimental parameters, it was necessary to define which parameters would be studied over what range.

Three main parameters were considered: scan rate V, number of cycles nCy, switch potential E_{sw} , initial potential E_i.

Note that effects of the sampling parameter, β, are not included in this study, because the DigiSim® software does not provide for this option. Thus, the simulated voltammograms here should be referred to more properly as cyclic voltammograms (CV) ratter than CSCV (sections 2.1 and 2.2). This terminology will be used below.

Because a fractional factorial design (see below) was used to define the boundaries of the study, voltammetric parameters were considered at two levels each. The high level scan rate is set to 1.00 v/s; while low level scan rate is set to 0.05 v/s. The high level number of cycles is set to two; while low level is set to one. The high level E_{sw} verse the peak potential, E_p, is equal to -300 mV; while low level is equal to -50 mV. In order that all the voltammograms have a fixed range (1280 mV), the corresponding E_i for high level is +980 mV, and low level is +1230 mV. Because fourier features computed by Microsoft® EXCEL fourier analysis has a 2^n (n = positive even integer number) data input range requirement, we have selected 1024 as the total number of simulated CV data points.

For a one-cycle CSCV voltammogram:

The step size ΔE = 0.0025 v (2.5 mV), giving a window for the CV of

ΔE * (data points/2) = 2.5 mV* (1024/2) = 1280 mV.

For a two cycle CSCV voltammogram, in order to have the same 1024 data points, the step size ΔE has to be twice that of the one cycle CSCV. So,

The step size ΔE = 0.005 v (5.0 mV), giving a window for the CV of

ΔE * (data points/4) = 5.0 mV* (1024/4) = 1280 mV.

Thus, both the one-cycle and two-cycle voltammograms examine the same voltage window, with the experimental cathodic E_p of cycle one always set arbitrarily a 0.000 V; and the peak potential always occurs at either 50 mV before E_{sw} or 300 mV before E_{sw} (low and high level).

3.3.3 Fractional Factorial Design

A factorial design approach (49, 50) was used to specify experimental parameters (Table 2) and simulation variables (compound properties) (Tables 3-6). Note that this study is only included two of the three variables specified in Table 2, scan rate, V, and number of cycles nC_y. In Tables 2-6, "-" refers to low level; "+" refers to high level.

According to DePalma and Perone's study (51), three kinetic parameters (the reversibility of the electron transfer, the symmetry factor α, and the number of electrons involved in the transfer, n) are used to define the shape of a voltammogram in the absence of the physical or chemical complications. The reversibility of electron transfer is dictated by the standard heterogeneous rate constant k_s.

In addition to the heterogeneous kinetic parameters, voltammogram shape is also influenced by the presence of coupled chemical or physical phenomena, viz., coupled chemical reactions or coupled surface adsorption. We consider only three types of coupled chemical processes in our work: succeeding equilibrium (EC), preceding equilibrium (CE) and the combined processes (ECE) (see section 3.3.1).

Additional voltammetric shape parameters include K, k_f (or k_b), for the homogeneous reaction. Thus, the variables ks, α, K, k_f , form a matrix of four factors with two levels (high, low) for a statistically designed set of representative voltammograms.

Table 3 defines a matrix of kinetic parameters corresponding to 25 different "compounds". Each "compound" has a number assigned indicated in the Table 3 matrix.

For the reversible case, ks = 10^6, α = 0.5; for quasi-reversible case, ks = 0.0005 (low), 0.05 (high), α = 0.3 (low), 0.7 (high). For both reversible and quasi-reversible cases, k_f low level is equal to 0.2, and k_f high level is equal to 10.

The values of K are varied from 0.005 to 100 depending on the mechanism (Table 4 and Table 6).

Table 5 is an expanded matrix where α and k_s are at three levels each. Also, for the EC case, the range of the homogeneous equilibrium constant K was reduced from 1000 to 50 by changing the high level magnitude from 50 to 2.5 (Table 6). As in Table 3, each "compound" has a number assigned, corresponding to 27 additional "compounds" in Table 5, for a total of 52 "compounds".

Table 2. Experimental parameters matrix. Entries are identifies labels for simulated voltammograms.

V (V/s)	E_{sw}/E_i (mV)	$n C_y^- = 1$	$n C_y^+ = 2$
$V^- =$ 0.05	$(E_{sw}/E_i)^-$	A	A2
	$(E_{sw}/E_i)^+$	(B)	(B2)
$V^+ =$ 1.00	$(E_{sw}/E_i)^-$	C	C2
	$(E_{sw}/E_i)^+$	(D)	(D2)

$(E_{sw}/E_i)^-$: $E_{sw} = -50$ mV, $E_i = 1230$ mV

$(E_{sw}/E_i)^+$: $E_{sw} = -300$ mV, $E_i = 980$mV

Table 3. Chemical parameters matrix I. Entries are identifying numbers for individual simulated voltammograms (see Table 4 for values of K, k_f).

	Reversible		Quasi-reversible									
	$k_s = 10^6$		$k_s = 0.0005$				$k_s = 0.05$					
	$\alpha = 0.5$		$\alpha = 0.3$		$\alpha = 0.7$		$\alpha = 0.3$		$\alpha = 0.7$			
	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+		
	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f		
	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +		
EC	6	9	12		15		18	21		24		
CE	7		10	13	16		19		22			
ECE	8		11		14	17	20		23	25		
E	1		2		3		4		5			

Table 4. Values of K, k_f for chemical parameters matrix I.

Mechanisms	K^-	K^+	$k_f^-(s^{-1})$	$k_f^+(s^{-1})$
EC	0.05	50	0.2	10
CE	0.005	0.1	0.2	10
ECE	1	100	0.2	10

Table 5. Chemical parameters matrix II. Entries are identifying numbers for individual simulated voltammograms (see Table 5 for values of K, kf).

	Reversible	Quasi-reversible																							
	$k_s = 10^6$	$k_s = 0.0005$						$k_s = 0.005$						$k_s = 0.05$											
	$\alpha = 0.5$	$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 0.7$		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 0.7$		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 0.7$							
	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	K^-	K^+	
	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	k_f	
	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	- +	
EC	6	9	12		77	80	15			83		86		89		92	18	21		95		24			
CE	7	48	10	13	78			16	81	84		87		90			19		93		22				
ECE	27	8	11			79	14	17	82		85	88		91				20		94	96	23	25		
E	1		2		71		3		72		73		74		4		76		5						

Table 6. Values of K, k_f for chemical parameters matrix II.

Mechanisms	K^-	K^+	$k_f^-(s^{-1})$	$k_f^+(s^{-1})$
EC	0.05	2.5	0.2	10
CE	0.005	0.1	0.2	10
ECE	1	100	0.2	10

3.4 Feature Definition and Extraction

Traditionally, cyclic voltammograms have been characterized by quantities like peak potentials, and peak current ratios etc., and their dependence on initial potential, switching potential, scan rate and other variables (21).

The approach used here, however, is to consider the fourier transform of each voltammogram as a concise, objective way of describing all shape factors. Dependencies of these measures on the traditional experimental parameters are obtained by conducting studies defined by the factorial matrix of Table 2.

In order to compare voltammograms, they must be collected under the same conditions, examine the same size voltage window, and be submitted to the same preprocessing.

First, in order to avoid introducing extraneous shape information into the transform, each curve must have the same peak location. This condition is accomplished by shifting the simulated voltammogram curve so that its peak potential equals zero ($< \pm 0.01$ mV).

Second, each curve is rotated and translated prior to the fourier transform to retain the most fidelity in the transform (52). Each voltammogram is rotated and translated in such a way that high frequency artifacts in the transform are avoided, so that the first and last data points are zero, and no discontinuities are introduced. The raw CV data file is translated/rotated by subtracting a linear term which is computed by connecting the first current value to the last. This method has been shown to

be an effective means of maintaining the fidelity of electrochemical data during fourier transformations (6).

Third, the fast fourier transform (FFT) is calculated by using the Microsoft® EXCEL analysis tool package, and only the real part of the complex number from the output data is used as fourier coefficients. The 1024 voltammetric data result in 512 fourier coefficients.

Finally, the fourier coefficients are normalized to the maximum absolute value. The 512 normalized magnitudes are used as fourier features, whose values range from -1.0 to 1.0. Those features are used for further pattern recognition study.

3.5 Pattern Recognition Procedure

The K-nearest neighbor (KNN) approach (section 2.4) has been used for achieving the goal of finding an optimum set of features which will accurately classify a set of voltammetric data.

The KNNPRA software used here requires that all the fourier transform data files are converted to text files before running the pattern recognition program. The KNNPRA software also limits the number of pattern descriptors to 17, of which the first two are reserved for pattern ID and class. This leaves space for the maximum of fifteen features. In the experiment, three fourier transform text files containing the first 45 features were examined separately by the KNNPRA software. The most useful features providing the highest class accuracy (section 2.4.3.2.1) from each of these three groups are combined into a new text file, and the pattern recognition program is run again.

In our study, KNN training with feature elimination and KNN training with class-weighted training were used for pattern recognition procedures.

KNN training with feature elimination procedure: the KNN program examines all patterns in a training set using a leave-one-out (LOO) K-nearest neighbor pattern classification algorithm. This procedure assigns a class to each pattern as if it were an "unknown" , after all patterns are classified by the LOO procedure, the overall accuracies are determined and printed out. The operator can select which of all possible features are used for KNN classification (operator selects "0" weights for unwanted features). Also, as described in section 2.4.4, forward or backward feature elimination can be selected by the operator. The operator can also select whether or not feature weight optimization occurs during the feature elimination procedure.

The end result of the training process is the identification of several combinations of features and weights providing the maximum training classification accuracy. The validity of the several feature sets deemed promising from training are then evaluated by application to prediction sets. Accurately input pattern/feature and other data base characteristics are required by the operator.

KNN Training with Class Weighted Training Procedure: this procedure is very similar to the above KNN training with feature elimination, except that the operator selects one class for which class-specific accuracy is optimized. Feature selection and weighting uses a class weighted classification error factor as figure of merit, where the error is:

$$\text{Error} = [\text{total \# wrong}] + [\text{\# wrong selected class}] + [\text{\# false positives}] \quad (14)$$

In our studies, pattern recognition class accuracy was evaluated one class at the time, selecting a different set of features which maximized the class specific accuracy for each class. Class-weighted training was used at some point in the overall training process, but the first step in the feature elimination procedure does not involve class weighting.

Chapter 4 Results and Discussion

4.1 Overview of these Studies

The body of this study involves three parts, summarized in Figure 11.

First, simulated cyclic voltammetric (CV) data were generated for hypothetical chemical compounds of five different classes (defined in section 3.3.1) Second, pattern recognition techniques were used to identify compounds/classes from fourier transform CV data and to determine which features and/or experimental parameters were most useful for accurate classifications. Third, information theory was used to provide a more quantitative measure of information content related to experimental parameters for qualitative identification of redox species.

4.2 Organization of Pattern Sets

Tables 2-6 define the range of properties of hypothetical compounds in the data base as well as the specific sets of voltammetric conditions considered. In order to examine separately the effects of various factors, several different combinations of patterns were studied by pattern recognition, obtaining results correlated with different factors. The different pattern sets are described in Tables 7 to 11, where N = number of compounds, P = number of patterns. Each "compound" corresponds to a specific set of

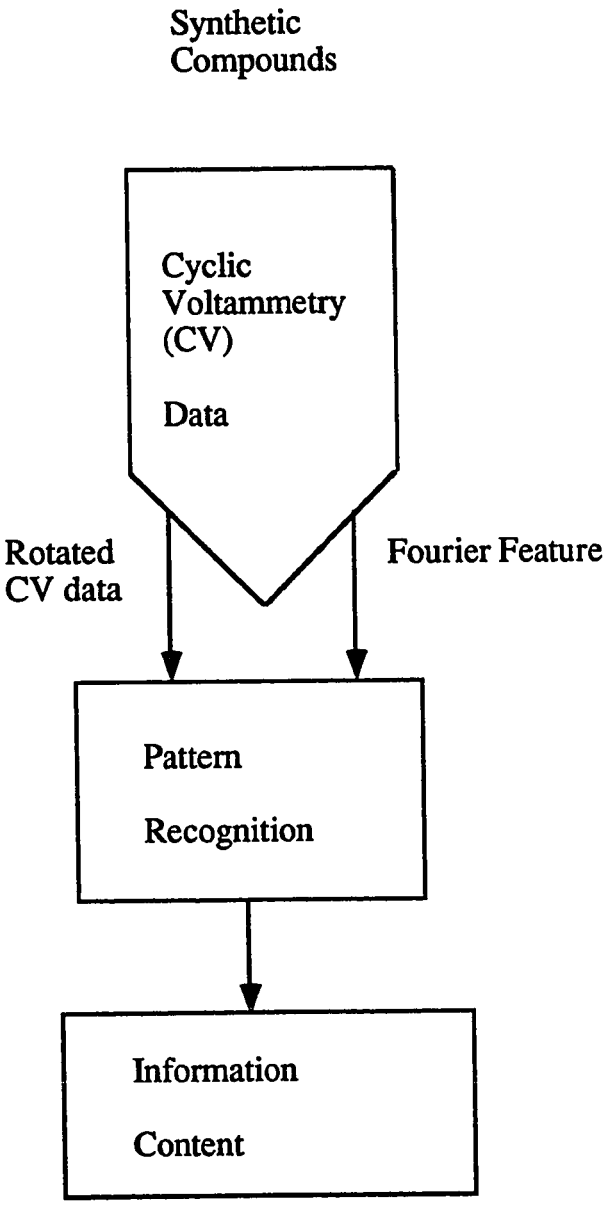

Figure 11. Procedure of the study.

Table 7. Baseline set A, N = 17, P = 34, experimental conditions A and C (Table 2).

Class	# of patterns	CV # (Table 3)
1	2	1
2	8	2 3 4 5
3	8	6 9, 15 21
4	8	7 10 19 25
5	8	8 14 17 23

Table 8. Pattern sets to study effects of set size and distribution.

a. N = 21, P = 42, experimental conditions A and C (Table 2), quasi-reversible conditions only for classes 2-5.

Class	# of patterns	CV # (Table 3)
1	2	1
2	8	2 3 4 5
3	10	12 15 18 21 24
4	12	10 13 16 19 22 25
5	10	11 14 17 20 23

b. N = 25, P = 50, experimental conditions A and C (Table 2).

Class	# of patterns	CV # (Table 3)
1	2	1
2	8	2 3 4 5
3	14	6 9 12 15 18 21 24
4	14	7 10 13 16 19 22 25
5	12	8 11 14 17 20 23

c. N = 52, P = 104, experimental conditions A and C (Table 2).

Class	# of patterns	CV # (Table 5)
1	2	1
2	18	2 3 4 5 71 72 73 74 76
3	28	6 9 12 15 18 21 24 77 80 83 86 89 92 95
4	28	7 10 13 16 19 22 25 48 78 81 84 87 90 93
5	28	8 11 14 17 20 23 27 79 82 85 88 91 94 96

Table 9. Pattern sets to study scan rate effects.

a. Same CV number's as Table 7, baseline A, N = 17, P = 18 (2 identical patterns for class 1), each pattern is obtained by subtracting low scan rate voltammograms from high scan rate voltammograms for the same "compound" and conditions.

Class	# of patterns	CV # (Table 3)
1	2	1
2	8	2 3 4 5
3	8	6 9, 15 21
4	8	7 10 19 25
5	8	8 14 17 23

b. Same CV number's as Table 8a, N = 21, P = 22 (2 identical patterns for class 1).

Class	# of patterns	CV # (Table 3)
1	2	1
2	8	2 3 4 5
3	10	12 15 18 21 24
4	12	10 13 16 19 22 25
5	10	11 14 17 20 23

c. Same CV number's as Table 8b, N = 25, P = 26 (2 identical patterns for class 1).

Class	# of patterns	CV # (Table 3)
1	2	1
2	8	2 3 4 5
3	14	6 9 12 15 18 21 24
4	14	7 10 13 16 19 22 25
5	12	8 11 14 17 20 23

Table 10 Baseline set B, N = 17, P = 34, experimental conditions A and C (Table 2).

Class	# of patterns	CV # (Table 3)
1	2	1
2	8	2 3 4 5
3	8	9 12 21 24
4	8	10 13 16 19
5	8	14 17 20 23

Table 11. Pattern sets to study effect of number of cycles, same CV number's as Table 10, N = 17, P = 34, experimental conditions A2 and C2 (Table 2).

Class	# of patterns	CV # (Table 3)
1	2	1
2	8	2 3 4 5
3	8	9 12 21 24
4	8	10 13 16 19
5	8	14 17 20 23

chemical/physical parameters as defined in Tables 3 and 5, and is assigned a unique ID number. Several patterns may be obtained for each "compound", using the various combinations of experimental parameter values defined in Table 2, and identified by symbol A, B, C, D and A2, B2, C2, D2.

Also, note that for all pattern sets, class 1 (the uncomplicated reversible case) is represented by only two patterns. This is because the CV shape for this class does not change with any of the parameters studied. However, at least two CV's are required so that each can be classified by comparison to one other member of the same class.

The purpose of the Baseline A pattern set of Table 7 is to establish a baseline of pattern recognition performance to compare with other sets obtained under different conditions.

Table 8 includes pattern sets to study effects of set size and distribution.

Table 9 includes pattern sets to study scan rate effects

The purpose of the Baseline B pattern set of Table 10 is to establish a baseline of pattern recognition performance to compare with subsequent pattern sets, where the number of cycles is two rather than one. Table 10 differs from Table 7 in that some different "compounds" (Table 3).are included.

Table 11 is the pattern set to study the effect of number of cycles.

In addition to varying the experimental parameters, changing procedural factors had an impact. For example, different subsets of Fourier transform features were selected for KNN training with feature elimination; this leads to the retention of different sets of useful features and varying classification Performance.

Besides different feature combinations, sequences followed for KNN training were studied. That is, the point at which class-weighted training started was varied. Specifically, two cases were studied. Using the pattern set of Table 8b, training was conducted as :

1). Normal KNN with feature elimination and weight optimization, followed by class-weighted feature elimination and weight optimization with the remaining features.

2). Normal KNN with one-pass feature elimination, followed by class weighted feature elimination and weight optimization with the remaining features.

4.3 Interpretation of Pattern Recognition Results

The definition of various combinations of patterns for pattern recognition studies (Section 2.4) allowed evaluation of the effects of various factors on voltammetric information content. The relevant informational goal in these studies is accuracy of identification of individual compound classes, utilizing only voltammetric shape features derived from fourier transform CV data. The effectiveness of various experimental factors was evaluated by computing information content from pattern recognition classification accuracy data. The basis of these calculations is described below.

4.3.1. Information Content

The information content for a particular qualitative identification exercise can be expressed in terms of the classification accuracy. However, this expression is not readily comparable to another situation where different numbers of samples or different class distributions are involved. Thus, we have attempted to use information theory principles (specifically, Shannon's theorem, Section 2.3.1) to express information content in terms which are more generally acceptable and comparable to other studies.

Applying Shannon's theorem to the 5-class problem considered here, we can define five possible outcomes of any given experiment as the identification of a sample as belonging to one of five classes. Furthermore, because we know the true classs of each sample, we can define each experiment as producing either one of five correct results, or no result. From Equation 2 of Section 2.3.1 we can describe the average information content of this classification exercise as:

$$H_{av} = - \sum_{R=1}^{z} (C_R/J)\log_2(C_R/J) \qquad (15)$$

where J = total no. patterns;

C_R = total correctly classified of class R;

Z = total no. classes,

If $C_1 = C_2 = \ldots\ldots C_5 = J/Z$, then $H_{max} = \log_2 5 = 2.32$ bits.

Because the class distributions in our study are not equal, the value of H_{max} is variable for each set. For example, for the class distribution of Table 7, with 100% class-specific classification accuracy, H_{max} = 2.20 bits. For Table 8c, H_{max} = 2.08 bits.

A more appropriate expression can also be derived from Shannon's theorem, describing the information gain obtained before and after pattern classification (53-56). The related expression is:

$$\Delta H = \log_2(n_o/n) \tag{16}$$

where n_o is the total number of possible identities that might be assigned to an "unknown" item, and n is the number of possible identities which might be assigned to an "unknown" item after the pattern recognition procedure. Ideally, $n = 1$ for perfectly accurate recognition. Thus, if n_o is 10 and n is 1, ΔH is 3.32 bits.

For the classification of a number of items, the average information gain is given by:

$$\Delta H_{av} = (1/J) \sum_{j=1}^{J} (\log_2(n_o/n)_j) \tag{17}$$

For the qualitative identification exercises considered in this study, where a five-class recognition problem is considered, if there are equal numbers of each class and n_o/n is 5, the information gain for each correctly identified item is $\log_2 5$, or 2.32 bits. In this study, there is no partial reduction of uncertainty by mis-classification. Thus, for an incorrectly identified item, the information gain is $\log_2(n_o/n_o)$, or $\Delta H = 0$.

Considering these definitions, the maximum information gain can be calculated from Equation 17 for the five-class identification problem of Table 12, where J=34, N=17, n_0 is 17 for all classes, and n_i is (1, 4, 4, 4, 4) for each accurately identified item in classes 1 to 5. Thus, for 100% accurate classification, ΔH_{mx} = 2.20 bits. For the best results of Table 12 the information gain was 1.71 bits (76.5% overall accuracy). (NOTE: Because of the way each pattern set is designed, class 1 accuracy is always 100%. Only two examples of this class are included in each pattern set to minimize impact on ΔH values. Moreover, class-specific accuracy for class 1 is not reported in the results tables).

Using this definition (Equation 17) for information gain it is possible to make direct comparisons between the results of different classification problems considered in this study, where differing class distributions are involved. In these cases, the classification accuracy figure of merit is inadequate. Moreover, the information gain figure of merit is more meaningful for interpretation of factor effects considered below.

4.4 Observations of Classification Accuracy and Information Content

Tables 12 to 23 summarize the results of various pattern recognition training studies. Classification performance is expressed in various ways. The features listed by number refer to the indices of the fourier transform coefficients. For the results of Tables 12 - 18, the pattern recognition

procedure did not involve class-weighted training (except for Table 14). Results of Tables 19 - 23 involved class-weighted training.

The effects of several factors on classification accuracy and/or information content were studied. These include data base properties, procedures, and experimental parameters. Variable data base properties include: numbers of "compounds" and patterns, distribution of "compound" classes, and distribution of "compound" characteristics. Variable procedures include: starting feature combinations and training sequence. Variable experimental parameters include: scan rate, number of cycles, and switching potential. Of all these parameters, only the effects of switching potential were not investigated in this study.

4.4.1. Effects of Numbers and Distributions of Patterns

The evaluations of effects of some factors on classification accuracy necessarily require changes in numbers of patterns in the training set. For example, see Tables 7, 8a, 8b, 8c, the use of difference patterns reduces the total number, P, by about one-half for the same number of "compounds", N. Thus, it was important to determine if the number of patterns in the training set had a significant effect on information content.

To answer this question, one can compare the classification results (best feature set, overall accuracy) obtained in Tables 12 and 13 with those from Tables 18, 19, and 20. Despite the increase in the number of patterns in the training set from 34, 42 to 50, the overall classification accuracy did not

improve. In fact, a slight decrease was observed (from an average of 73.5% to 65.3%). The related average information gain values are 1.65 (±0.06) and 1.47 (±0.02) bits, respectively. These are 75.0% and 69.0% of ΔH_{mx}. Although small, this negative effect of training set size on information gain is significant.

These observations suggest that for subsequent evaluations of other factors, the effects of incidental changes in numbers of patterns on information content should be considered.

The training sets for Tables 12 and 18 represent the same distribution of voltammetric parameters. However, when the results of Tables 20 and 21 are compared, not only is the number of patterns increased, but there is a change in the distribution of CV parameters (Table 3 vs Table 5 parameters for the training sets of Tables 20 and 21, respectively). The result is that the information content for Table 21 training sets is greater than for Table 20 (For the best feature sets, overall accuracies are 68.0% and 74.0%, respectively). The average information gains (for the best overall accuracy) are 1.50 and 1.57 bits for Tables 20 and 21. This increase is significant, particularly considering the expected negative effect due to the increased training set size from Table 20 to 21.

4.4.2 Procedural Effects

4.4.2.1 Effects of Training Feature Combinations

Because the training procedure followed in this study involved examining 15 features at a time, and 50 total fourier coefficient features were

examined for each voltammogram, it was important to determine whether there was any effect of the sequence in which features were examined on training results. This concern was examined in several studies, with the conclusion that there was no significant effect. A good example of a comparative study is illustrated in Tables 12 and 13. The pattern classification results are reported for the same pattern set (Baseline A set, Table 7). For Table 12 study the features were divided into four sequentially indexed groups; for Table 13 the features were divided into various size sub-sets which were inter-leaved for training. The best overall accuracies achieved are 76.5 and 73.5%, respectively. (Note that only the best of many possible starting combinations are listed in Table 13).

Because of this observation, most training studies were conducted by examining the features according to the sequence indicated in Table 12. This was convenient, and there was no indication that other procedures would produce significantly different results.

4.4.2.2 Effects of Class-Weighted Training Sequence

When class-weighted training was applied (see section 3.5), two different training sequences were considered. These were:

(A) apply class-weighted with weight optimization training to the feature sub-set remaining after completing the normal systematic trial-and-error feature elimination procedure for as many cycles as required to condense the feature set to a minim set with maximum accuracy.

and

(B) apply class-weighted with weight optimization training to the feature sub-set remaining after only one forward and backward feature elimination cycle.

The classification training results of Tables 19 and 20 reflect the application of procedures **A** and **B**, respectively. These tables report class-specific accuracies for various feature sets. This classification procedure finds optimum features for identifying one class at a time. These binary decisions can produce a ΔH_{mx} of 1.00 bit for two classes of equal size. In this study ΔH_{mx} varied for each class. For example, for class 2, ΔH_{mx} = 0.63 bits; for classes 3 and 4, ΔH_{mx} = 0.86; for class 5 ΔH_{mx} = 0.80. The average class-specific information gain for the best feature sets is 0.63 for both Tables 19 and 20. The average relative information gain ($\Delta H/\Delta H_{mx}$) is 79.0% (±12.0) and 75.4% (±13.5), respectively. Clearly, there is no significant difference in training accuracy with either procedure. Because it was more convenient, procedure **B** was used for all other studies involving class-weighted training.

4.4.3 Effects of Experimental Parameters

4.4.3.1 Voltage Scan Rate Factor

Traditionally, scan rate, V, has been the experimental variable most commonly used to characterize electrode processes voltammetrically. Thus, in our study the effect of this variable was examined in connection with the

study of all other experimental parameters. For Tables 16 and 17, two pattern subsets were used for training. One included CV's at low scan rate (A); the other included only CV's at high scan rate (C).

For results of Tables 16 and 17, the classification accuracies are compared for the low V and high V pattern sub-sets. Clearly, low scan rate provides greater classification accuracy (72.7% VS. 45.5% for the best results of Table 16; 65.4% VS. 57.7% for Table 17). The maximum of information gain is 1.71 bits (72.7% overall accuracy).

4.4.3.2 Voltage Scan Rate Effect

A more direct way to examine the effect of scan rate on classification performance is to compare training results for the same set of "compounds", first with all voltammograms collected at two different scan rates, then with a pattern set composed of difference voltammograms. That is, each pattern is obtained by taking the difference between the high-scan and low-scan rate voltammograms for a given "compound", with all other parameters being equal.

To evaluate the scan rate effect, one can compare the classification results obtained from Tables 12 and 14. Table 14 training set included difference voltammograms obtained from the data in Table 12 training set. For the best results in Table 12, ΔH = 1.71 bits (76.5% overall accuracy), for the best results in Table 14, ΔH = 1.73 bits (72.2% overall accuracy). It is important to remember that the pattern sets compared are of significantly different size;

and earlier discussion demonstrated that information gain should decrease with increasing training set size. Thus, if there were no scan rate effect, it would be expected that the information gain would increase for difference voltammograms due to the effect of decreased training set size. For two sets of Table 12 and 14, the value of ΔH remains nearly constant. This suggests a possible negative effect on ΔH for difference voltammograms (scan rate effect).

4.4.3.3 Effects of Number of Cycles

The effect of using two-cycle voltammograms is compared to using single-cycle voltammograms in the results of Tables 22 and 23. The training studies lead to very similar results, with the information content of the two-cycle voltammograms coming out slightly higher (0.67 vs 0.68 information bits). Thus, there does not appear to be any significant advantage to using two-cycle CV's. This observation is surprising because of the intuitive perception that more information should come from two cycles. Clearly, these studies indicate otherwise, at least for the kinds of processes studied here.

4.5 General Observations

Class-specific training has identified some very useful classifiers from fourier transform voltammograms. Individual classes of compounds can be identified with 87.5% (class 2, 3) 100% (class 4), and 89.3% (class 5) accuracy.

The effects of procedural factors have been evaluated, and there appear to be significant effects of both the training set size (negative effect) and distribution (positive effect) on information gain. Of the experimental parameters studied here, the results of number of cycles studied are the most significant. That is, contrary to what was expected, there is no significant enhancement of information gain for two cycles compared to one. Surprisingly, there appears to be a negative effect when scan rate difference voltammograms are used for training set patterns. The information gain observed for training sets containing only high scan rate voltammograms was decreased compared to those using only low scan rate voltammograms.

It is tempting to speculate on the implication of these observations. Perhaps the easiest observation to explain is the positive effect of enhance pattern distribution, because it is expected that multivariate classification procedures should be more successful when there is a more nearly continuous distribution of pattern vectors for each class.

The negative effect of increased training set size is puzzling. This effect may be fortuitous, given the relatively small sets which were studied, or it might reflect the fact that class assignments are too arbitrary. Perhaps a factor analysis study would be useful to resolve this issue.

The enhanced information content observed for lower scan rate probably reflects the fact that for the particular electrode processes studied the effects of coupled chemical reactions are more pronounced at slower scan rate, allowing better discrimination among classes from voltammetric shape data.

The fact that increasing number of cycles does not increase information content is probably related to the fact that, for these particular

electrode processes, no new redox couple is observed during the second cycle. Thus, no new information is introduced by the second cycle. Conversely, for electrode processes where succeeding reactions lead to new redox couples (e.g. where the subsequent chemical product is reversibly oxidized at a potential much more positive of E^0 for the main redox couple) added information should be obtained for cycle two.

Regarding the specific fourier coefficient features which were useful for class identification, there were far more combinations which could accurately classify class 4 than for any other class. Features 1, 4 and 9 were the most commonly observed useful features.

It is important to note that these observations are based on a very limited data set, and that we were investigating **only** the information content of voltammetric **shape** data. The maximum information gain observed in any of the studies was about 1.75 bits. To put this in perspective, 2.32 bits would be required for 100% accurate classification into each of five classes. Moreover, to accurately identify each of 50 compounds would represent an information gain of 5.32 bits. This amount of information may not be obtainable from **shape** information alone. To enhance the total information content of the data one can simply add the peak potential data. Assuming that E_p data can be obtained with 5 mV resolution over a 1.500 V range, the calculated $\Delta H_{mx} = \log_2 300 = 8.23$ bits.

Table 12. Pattern classification results for Baseline A pattern set (Table 7), overall accuracy information gain ΔH_{mx} = 2.206 bits.

Fourier Feature Set	Best Features	Overall Accuracy %	Info. Gain (bits)
1 to 15	1 10 12	76.5%	1.687
16 to 30	17 20	61.8%	1.363
31 to 45	34 38 40 or 34 37 41	73.5%	1.621
46 to 50	46,49 or 46 47	61.8%	1.363
1 to 50	12 38 46 or 12 46 47	70.6%	1.557

Table 13. Pattern classification results for Baseline A pattern set (Table 7), overall accuracy information gain ΔH_{mx} = 2.206 bits.

Fourier Feature Set	Best Features	Overall Accuracy %	Info. Gain (bits)
1 to 7,15 to 21	2 6 17 18 19	67.6%	1.491
8 to 21	8 13	73.5%	1.621
8 to14, 22 to 28	8 13	73.5%	1.621
1 to 7, 22 to 28	7 22	67.6%	1.491
1 to 7, 29 to 35	6 35	67.6%	1.491
8 to14, 29 to 35	8 12 13 35	73.5%	1.621

Table 14. Pattern classification results for scan rate effect pattern set (Table 9a), overall accuracy information gain ΔH_{mx} = 2.206 bits, Class specific accuracy information gain for all Class 2, 3, 4, 5 are ΔH_{mx} = 0.787.

Fourier Feature Set	Class	Best Features	Class Acc. %	Overall Acc. %	FP	Info. Gain (bits)	Ratio
1 to 8	2	4 8	50.0%	61.1%	1	0.530	0.673
	3	6	50.0%	61.1%	2	0.519	0.659
	4	1 4	100.0%	56.7%	1	0.776	0.986
	5	4 8	50.0%	61.1%	1	0.530	0.673
9 to16	2	11 12 14	75.0%	51.0%	2	0.642	0.816
	3	12 14 16	50.0%	55.6%	3	0.507	0.644
	4	12 14	75.0%	44.4%	0	0.664	0.844
	5	12 14 15	25.0%	44.4%	0	0.419	0.532
17 to 24	2	19	50.0%	55.6%	4	0.496	0.630
	3	19	50.0%	55.6%	1	0.530	0.673
	4	19	75.0%	55.6%	1	0.653	0.830
	5	19	25.0%	55.6%	0	0.419	0.532
25 to 32	2	26 32	75.0%	33.3%	3	0.630	0.801
	3	-	0.0%	0.0%	-	0.000	0.000
	4	26 32	25.0%	27.8%	0	0.419	0.532
	5	-	0.0%	0.0%	-	0.000	0.000
33 to 40	2	36 39	75.0%	66.7%	2	0.642	0.816
	3	36 39	50.0%	66.7%	2	0.519	0.659
	4	36 39	100.0%	66.7%	0	0.787	1.000
	5	36 39	25.0%	66.7%	0	0.419	0.532
41 to 48	2	46 48	75.0%	72.2%	2	0.642	0.816
	3	46 48	50.0%	72.2%	1	0.530	0.673
	4	46 48	100.0%	72.2%	0	0.787	1.000

	5	46 48	50.0%	72.2%	0	0.542	0.689
1 to 48	2	12 46	75.0%	72.2%	1	0.630	0.801
	3	14 16 46	50.0%	72.2%	1	0.530	0.673
	4	36 39	100.0%	66.7%	0	0.787	1.000
	5	4 12	50.0%	66.7%	0	0.542	0.689

Table 15. Pattern classification results for quasi-reversible pattern sub-set (Table 8a), overall accuracy information gain $\Delta H_{mx} = 2.167$ bits.

Fourier Feature Set	Best Features	Overall Accuracy %	Info. Gain (bits)
1 to 15	4 6 13 or 2 4 13	54.8%	1.188
16 to 30	25 28 29	57.1%	1.237
31 to 45	35 41 44	54.8%	1.188
46 to 50	48 or 46 47 49	45.2%	0.979

Table 16. Pattern classification results for quasi-reversible pattern sub-set, scan rate effect, (Table 9b), set A (low scan rate), set C (high scan rate), overall accuracy information gain ΔH_{mx} = 2.330 bits.

Fourier Feature Set (A)	Best Features	Overall Accuracy %	Info. Gain (bits)
1 to 10	5 10 or 1 4 10	50.0%	1.165
11 to 20	12 15 16	45.5%	1.060
21 ro 30	28 29	72.7%	1.694
31 to 40	1 10 or 1 3 10	68.2%	1.589
41 to 50	48 50	63.6%	1.482
Set (C)			
1 to 10	2 4 9 or 2 4 10	40.9%	0.953
11 to 20	11	45.5%	1.060
21 to 30	24 29 or 24 25 26	40.9%	0.953
31 to 40	33 35 40	40.9%	0.953
41 to 50	49 or 44 45 46	45.5%	1.060

Table 17. Pattern classification results for quasi-reversible pattern sub-set, scan rate effect, (Table 9c), set A (low scan rate), set C (high scan rate), overall accuracy information gain ΔH_{mx} = 2.280 bits.

Fourier Feature Set (A)	Best Feature	Overall Accuracy %	Info. Gain (bits)
1 to 12	2 10	65.4%	1.491
13 to 24	13 16	46.3%	1.056
25 to 36	25 29	61.5%	1.402
37 to 48	39 48 or 46 48	61.5%	1.402

Table 18. Pattern classification results for Baseline A distribution with larger pattern set (Table 8b), overall accuracy information gain ΔH_{mx} = 2.131 bits.

Fourier Feature Set	Bset Features	Overall Accuracy %	Info. Gain (bits)
1 to 15	2 13 14	64.0%	1.364
16 to 30	18 21 28	50.0%	1.066
31 to 45	34 38	60.0%	1.279
46 to 50	46,49 or 46 47	61.8%	1.317
1 to 50	2 38 or 2 13 14	64.0%	1.364

Table 19. Pattern classification results for Baseline A distribution with larger pattern set (Table 8b), overall accuracy information gain ΔH_{mx} = 2.131 bits. Class 2 accuracy information gain ΔH_{mx} = 0.634 bits,Class 3, 4 accuracy information gain ΔH_{mx} = 0.856 bits, Class 5 accuracy information gain ΔH_{mx} = 0.795 bits. FP = False Possibility.

Fourier Feature Set	Class	Best Features	Class Acc. %	Overall Acc. %	FP	Info. Gain (bits)	Ratio
1 to 15	2	2 13 14	50.0%	64.0%	4	0.403	0.636
	3	2 13	78.6%	54.0%	4	0.708	0.827
	4	2 13 14	78.6%	64.0%	1	0.736	0.86
	5	2 13 14	66.7%	64.0%	4	0.599	0.753
16 to 30	2	18 21 28	37.5%	48.0%	6	0.340	0.536
	3	18 21 28	28.6%	50.0%	9	0.403	0.471
	4	18 21 28	78.6%	50.0%	2	0.726	0.848
	5	18 21 28	50.0%	50.0%	4	0.516	0.649
31 to 45	2	34 38	12.5%	60.0%	8	0.234	0.369
	3	34 38	50.0%	60.0%	8	0.523	0.611
	4	34 38	85.7%	60.0%	0	0.782	0.914
	5	34 38	66.7%	60.0%	2	0.615	0.774

1 to 45	2	2 13 14	50.0%	64.0%	4	0.403	0.636
	3	2 38	78.6%	54.0%	4	0.708	0.827
	4	2 38	92.9%	64.0%	2	0.800	0.935
	5	2 38	66.7%	64.0%	3	0.607	0.764

Table 20. Pattern classification results for Baseline A distribution with larger pattern set (Table 8b), overall accuracy information gain ΔH_{mx} = 2.131 bits. Class 2 accuracy information gain ΔH_{mx} = 0.634 bits,Class 3, 4 accuracy information gain ΔH_{mx} = 0.856 bits, Class 5 accuracy information gain ΔH_{mx} = 0.795 bits. FP = False Possibility.

Fourier Feature Set	Class	Best Features	Class Acc. %	Overall Acc. %	FP	Info. Gain (bits)	Ratio
1 to 15	2	2 4 13	50.0%	58.0%	3	0.408	0.644
	3	2 4 13	57.1%	60.0%	7	0.569	0.665
	4	2 4 6 13	78.6%	58.0%	3	0.717	0.838
	5	2 4 6 13 14 15	50.0%	60.0%	3	0.524	0.659
16 to 30	2	25 28	37.5%	44.0%	6	0.340	0.536
	3	18 21 24 28 29	28.6%	46.0%	9	0.403	0.636
	4	22 24 29	92.9%	44.0%	1	0.809	0.945
	5	18 21 25 29	50.0%	50.0%	4	0.516	0.649
31 to 45	2	34 38 44	25.0%	58.0%	7	0.282	0.445
	3	38 44	42.9%	54.0%	9	0.476	0.556
	4	34 38	85.7%	60.0%	1	0.772	0.902
	5	34 38	66.7%	60.0%	2	0.615	0.774

1 to 45	2	2 13 44	50.0%	54.0%	4	0.403	0.636
	3	2 13 44	57.1%	68.0%	7	0.569	0.665
	4	4 13 38	85.7%	60.0%	1	0.772	0.902
	5	2 4 6 14	58.3%	56.0%	4	0.558	0.702

Table 21. Pattern classification results for Baseline A distribution with larger pattern set (Table 8c), overall accuracy information gain ΔH_{mx} = 2.077 bits. Class 2 accuracy information gain ΔH_{mx} = 1.227 bits, Class 3, 4 5 accuracy information gain ΔH_{mx} = 1.509. FP = False Possibility.

Fourier Feature Set	Class	Best Features	Class Acc. %	Overall Acc. %	FP	Info. Gain (bits)	Ratio
1 to 15	2	1 2 4	50.0%	63.5%	5	0.714	0.582
	3	1 4	64.3%	61.5%	11	1.182	0.783
	4	1 4 11 15	75.0%	56.7%	2	1.355	0.898
	5	1 2 4	82.1%	65.4%	4	1.365	0.905
16 to 30	2	24 28 29	22.2%	51.0%	18	0.401	0.327
	3	20 21 24 26 30	32.1%	48.1%	17	0.939	0.622
	4	20 21 24 26 29 30	82.1%	52.9%	2	1.391	0.922
	5	20 21 24 25 29	64.3%	51.0%	0	1.327	0.879
31 to 45	2	34 36 39	16.7%	51.0%	13	0.359	0.293
	3	33 34 39	32.1%	51.9%	8	1.058	0.701
	4	34 35 39 40	82.1%	52.9%	0	1.418	0.940
	5	34 39 45	67.9%	51.0%	11	1.201	0.796

1 to 45	2	1 2 34 39	66.7%	68.3%	8	0.872	0.711
	3	1 4	64.3%	61.5%	11	1.182	0.783
	4	1 23 30 39	85.7%	65.7%	0	1.436	0.952
	5	1 2 4 45	89.3%	74.0%	4	1.402	0.929

Table 22. Pattern classification results for two-cycle voltammograms (Table 11) overall accuracy information gain ΔH_{mx} = 2.206 bits. Class specific accuracy information gain for all Class 2, 3, 4, 5 are ΔH_{mx} = 0.787.

Fourier Feature Set	Class	Best Features	Class Acc. %	Overall Acc. %	FP	Info. Gain (bits)	Ratio
1 to 15	2	1 4 9	87.5%	67.6%	6	0.658	0.836
	3	4 9	87.5%	50.0%	0	0.726	0.922
	4	7 13	75.0%	55.9%	0	0.664	0.844
	5	3 4 7 8 13	75.0%	64.7%	2	0.642	0.816
16 to 30	2	25 28 30	62.5%	67.6%	4	0.557	0.708
	3	24 25	75.0%	50.0%	0	0.654	0.831
	4	25 29 30	87.5%	58.8%	0	0.726	0.922
	5	17 24 29	62.5%	58.8%	4	0.557	0.708
31 to 45	2	36 37 45	50.0%	41.2%	5	0.507	0.644
	3	35 37 38 42 25	75.0%	61.8%	0	0.726	0.922
	4	35 37 38 45	75.0%	50.0%	1	0.714	0.907
	5	39 42	50.0%	47.1%	3	0.507	0.644
1 to 45	2	1 28 37	75.0%	67.6%	3	0.630	0.801
	3	4 25 37 42	87.5%	58.8%	0	0.726	0.922
	4	25 35 37 38	87.5%	50.0%	0	0.726	0.922
	5	3 4 8 13	75.0%	70.6%	2	0.642	0.816

Table 23. Pattern classification results for Baseline B pattern set (Table 10), overall accuracy information gain ΔH_{mx} = 2.206 bits. Class specific accuracy information gain for all Class 2, 3, 4, 5 are ΔH_{mx} = 0.787.

Fourier Feature Set	Class	Best Features	Class Acc. %	Overall Acc. %	FP	Info. Gain (bits)	Ratio
1 to 15	2	2 11	75.0%	61.8%	1	0.714	0.907
	3	2 7 10 11	62.5%	64.7%	5	0.569	0.723
	4	2 7 11	75.0%	61.8%	1	0.714	0.907
	5	2 7 10	62.5%	61.8%	3	0.569	0.723
16 to 30	2	25 30	37.5%	52.9%	4	0.435	0.553
	3	28 30	50.0%	55.9%	1	0.530	0.673
	4	25 27 28	100.0%	55.9%	3	0.753	0.957
	5	27 28	62.5%	55.9%	3	0.569	0.723
31 to 45	2	34 40	37.5%	58.8%	5	0.446	0.567
	3	34 40	50.0%	58.8%	3	0.507	0.644
	4	34 41	87.5%	58.8%	0	0.726	0.922
	5	34 41	62.5%	58.8%	3	0.569	0.723
1 to 45	2	2 11	75.0%	61.8%	1	0.714	0.907
	3	7 10 40	75.0%	67.6%	4	0.692	0.879
	4	11 28	100.0%	52.9%	0	0.787	1.000
	5	2 7 10 27 28	75.0%	64.7%	3	0.630	0.801

Chapter 5 Conclusions

This study has provided some useful insight to the various parameters which might influence the ability to obtain qualitative identification information from fourier transform voltammetric data. It is clear that larger values for information content must be obtained for shape information alone to be very useful; but it is useful for some limited classification problems (small numbers of classes). The examination of the magnitude of classification problems that can be addressed when peak location information is added to the data base should be explored.

In addition, it would be worthwhile to examine the effects of other experimental parameters, like switching potential. The indications of effects of parameters already considered should be investigated further. For example, it is clear that an improvement in information gain can be achieved by improving the distribution of voltammetric properties. It is also clear that some electrode processes are more easily identified than others; so it would be worthwhile to extend the types of "compound" classes included in the data base.

Perhaps the most important achievement of this study was to lay the foundation for future studies of voltammetric information content. The quantitative treatment of information gain and the attention to experimental design will improve our ability to conduct future studies.

Chapter 6 Future Study

Considering the limited data set used, the results obtained here are very encouraging.

Further studies of the information content of voltammetric shape factors are warranted, utilizing a more extensive data base. In particular, the definition of larger numbers of compound classes should be done. And the promising positive effects of increased pattern distribution and use of lower scan rate conditions should be pursued. In addition, the effects of switching potential should be examined. Finally, it would be worthwhile to examine alternative multivariate analysis methods such as factor analysis to evaluate inherent information content.

Reference

1. S. P. Perone, and C. L. Ham, J. Research of the National Bureau of Standards, 90, number 6, 531, (1985).

2. C. E. Shannon, Bell System Technical Journal, 27,379, and 623 (1948).

3. P. A. Boudreau, and S. P. Perone, Anal. Chem., 51, 811, (1979).

4. Q. V. Thomas, and S. P. Perone, Anal. Chem., 49, 1369, (1977).

5. Q. V. Thomas, and S. P. Perone, Anal. Chem., 49, 1376, (1977).

6. D. R. Burgard, and S. P. Perone, Anal. Chem., 50, 1366, (1978).

7. W. A. Byers, and S. P. Perone, Anal. Chem., 52, 2173, (1980).

8. S. D. Schachterle, and S. P. Perone, Anal. Chem., 53, 1672, (1980).

9. W. A. Byers, and S. P. Perone, Anal. Chem., 55, 615, (1983).

10. W. A. Byers, and S. P. Perone, Anal. Chem., 55, 620, (1983).

11. A. J. Bard, L. R. Faulkner, Electrochemical Methods : Fundamentals and Applications, John Wiley and Sons: New York, 213-248, (1980).

12. L. L. Miaw, and S. P. Perone, Anal. Chem., 51, 1645, (1979).

13. S. S. Fratoni, and S. P. Perone, Anal. Chem., 48, 287, (1976).

14. K. F. Dahnnke, S. S. Fratoni, and S. P. Perone, Anal. Chem., 48, 296, (1976).

15. G. C. Baker, Advances in Polarography, I. S. Longmuir, Pergamon Press, New York, 144, (1960).

16. J. H. Christie, and P. J. Lingane, J. Electroanal. Chem., 10, 176, (1965).

17. J. J. Zipper, and S. P. Perone, Anal. Chem., 45, 452, (1973).

18. L. L. Miaw, P. A. Boudreau, M. A. Pichler, and S. P. Perone, Anal. Chem., 50, 1988, (1978).

19. C. K. Mann, Anal. Chem., 33, 1484, (1961), 35, 326, (1965), 36, 2424, (1966).

20. R. Bilewicz, R. A. Osteryoung, and J. Osteryoung, Anal. Chem., 58, 2761, (1986).

21. R. S. Nicholson, and I. Shain, Anal. Chem., 36, 706, (1964).

22. D. R. Ferrier, and R. R. Schroeder, Electroanal. Chem. and Interf. Electrochem., 45, 343, (1973).

23. R. R. Schroeder, and K. Lams, 169th ACS National Meeting, Philadephia, PA, April, (1975).

24. M. Ryan, J. Electroanal. Chem., 79, 105, (1977).

25. P. T. Kissinger, and W. R. Heineman, Laboratory Techniques in Electroanalytical Chemistry, Marcel Dekker: New York, 86-93, (1984).

26. J. W. Dillard, J. J. Odea, and R. A. Osteryoung, Anal. Chem., 51, 115, (1979).

27. J. W. Dillard, J. J. Odea, and R. A. Osteryoung, Anal. Chem., 51, 120, (1979).

28. T. Joslin, and D. Pletcher, J. Electroanal. Chem., 49, 171, (1974).

29. C. P. Andrieux, P, Hapiot, and J, Saveant, Chem. Rev., 90, 723, (1990).

30. C. P. Andrieux, P, Hapiot, and J, Saveant, Chem. Rev., 90, 739, (1990).

31. S.W. Feldberg in A. J. Bard (Ed.), Electroanalytical Chemistry, Vol. 3, Marcel Dekker, New York, 199, (1969).

32. M. Rudolph, D. P. Reddy, and S. W. Feldberg, Anal. Chem., 66, 589, (1994).

33. M. Rudolph, D. P. Reddy and S. W. Feldberg, DigiSim instruction tutorial.

34. C. E. Shannon, and W. Weaver, The Mathematical Theory of Communication, The University of Illinois Press: Urbana, 7-10, and 108-110, (1949).

35. S. P. Perone, Chemometrics, San Jose State University, (1992).

36. K. Eckschlager, and V. Stepanek, Anal. Chem. 54, 1115A, (1982).

37. H. C. Andrew, Introduction to Mathematical Techniques in Pattern Recognition, Wiley-Interscience, New York, (1972).

38. P. C. Jurs, and T. L. Isenhour, Chemical Applications of Pattern Recognition, Wiley-Interscience, New York, (1975).

39. M. A. Sharaf , D. L. Illman, and B.R. Kowalski, Chemometrics, Chemical Analysis, Vol. 82, Wiley-Interscience, New York, (1986).

40. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michette, and L. Kaufman, Chemometrics, a text book, Data Handling in Science and Technology, Vol. 2, Elsevier, Amsterdam, (1988).

41. D. R. Burgard, Chemometrics: Chemical and Sensory Data, (1990).

42. S. J. Haswell, Practical Guide to Chemometrics, (1992).

43. W. A. Whitney, IEEE Conference Record of the Symposium on Feature Extraction and Selection in Pattern Recognition, 49, (1970).

44. J. W. Sammon, IEEE Trans. Comput., C-18, 401, (1969).

45. K. Fukunaga, Introduction to Statistical Pattern Recognition, Chap.10, Academic Press, New York, (1972).

46. R. J. O'Halloran, and D. E. Smith, Anal. Chem., 50, 9, (1978).

47. P. R. Bevington, Data Reduction and Error Analysis for the Physical Science, Mcgraw-Hill, New York, (1969).

48. S. R. Schmidt, and R. G. Launsby, Understanding Industrial Designed Experiments, 3rd edition, Air Academy Press, Colorads Springs, (1992).

49. C. D. Hendrix, CHEMTECH, 9, 167, (1979).

50. G. E. P. Box, W.G. Hunter, and J. S. Hunter, Statistics for Experiments, Wiley, New York, Chap. 12, (1978).

51. R. A. DePalma, and S. P. Perone, Anal. Chem., 51, 829, 1979.

52. J. W. Hayes, D. E. Clover and D. E. Simth, Anal. Chem., 45, 277, (1973).

53. K. Eckschlager, and V. Stepanek, Information Theory as Applied to Chemical Analysis, Wiley, New York, (1979).

54. K. Eckschlager, and V. Stepanek, Analytical Measurement and Information, wiley, New York, (1985).

55. K. Eckschlager, and V. Stepanek, Anal. Chem., 54, 1115A, (1982).

56. L. Brillouin, Science and Information Theory, Academic Press, New York, (1962).